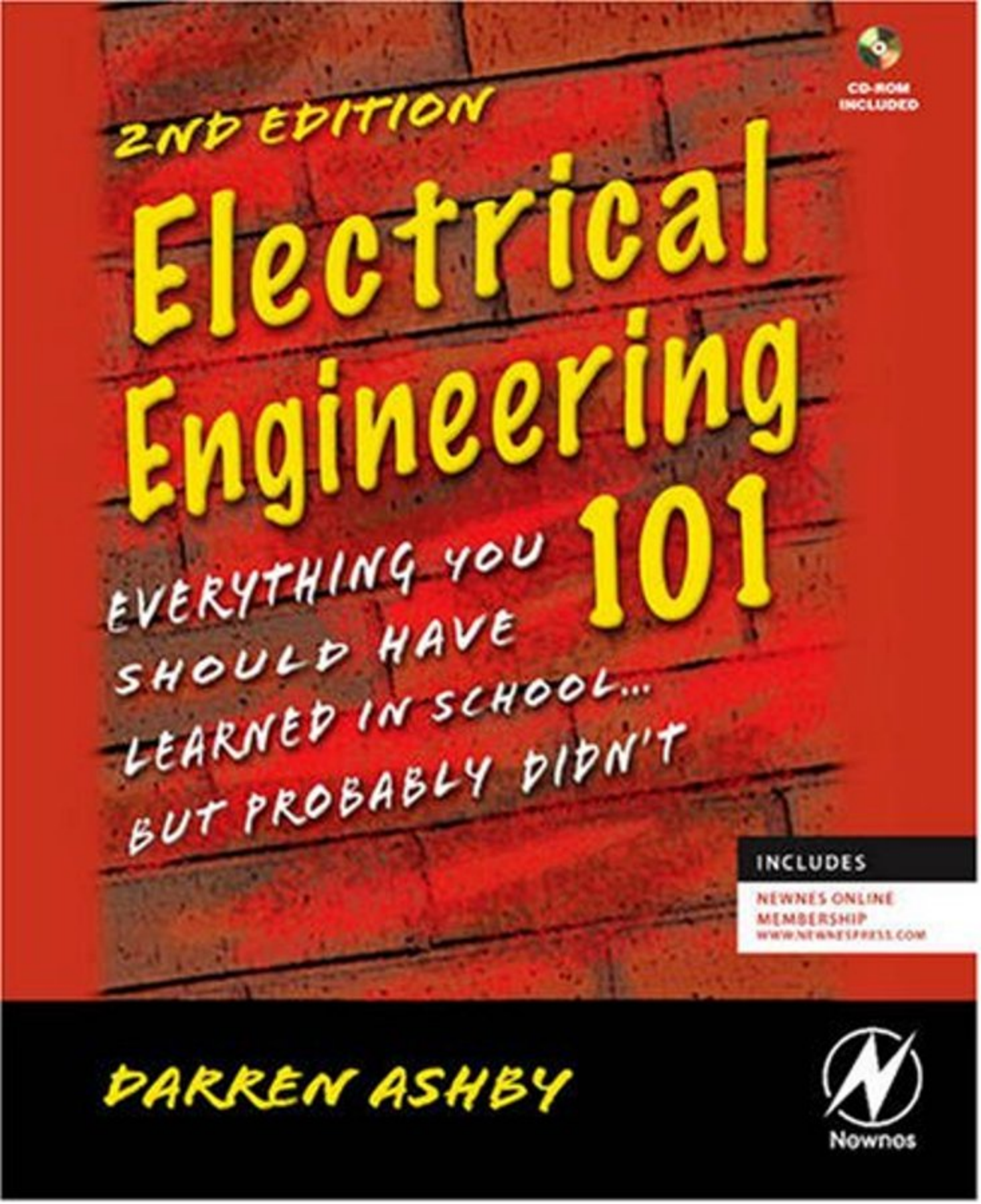

CD-ROM INCLUDED
2ND EDITION
Electrical Engineering 101
EVERYTHING YOU SHOULD HAVE LEARNED IN SCHOOL... BUT PROBABLY DIDN'T
INCLUDES
NEWNES ONLINE MEMBERSHIP
WWW.NEWNESPRESS.COM
DARREN ASHBY
Nownos

VISIT…

Newnes is an imprint of Elsevier
30 Corporate Drive, Suite 400
Burlington, MA 01803, USA

Linacre House, Jordan Hill
Oxford OX2 8DP, UK

Copyright © 2009, Elsevier Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, e-mail: permissions@elsevier.com.uk. You may also complete your request online via the Elsevier homepage (www.elsevier.com), by selecting "Customer Support" and then "Obtaining Permissions."

Recognizing the importance of preserving what has been written, Elsevier prints its books on acid-free paper whenever possible.

Library of Congress Cataloging-in-Publication Data

Ashby, Darren.
Electrical engineering 101 : everything you should have learned in school . . . but probably didn't / Darren Ashby.
p. cm.
Includes index.
ISBN 978-1-85617-506-7 (alk. paper)
1. Electric engineering. I. Title.
TK146.A75 2009
621.3—dc22

2008045182

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN-13: 978-1-85617-506-7

For information on all Newnes publications
visit our website at *www.books.elsevier.com*.

08 09 10 11 12 10 9 8 7 6 5 4 3 2 1

Printed in Canada

Working together to grow
libraries in developing countries
www.elsevier.com | www.bookaid.org | www.sabre.org
ELSEVIER BOOK AID International Sabre Foundation

Preface

THE FIRST WORD

Wow, the success of the original edition of *Electrical Engineering 101* has been amazing. I have had fans from all over the world comment on it and how the book has helped them. The response has been all I ever hoped for—so much so that I get a chance to add to it and make an even better version.

Of course, these days you don't just get a second edition, you get a *better* edition. This time through, you will get more insight into the topics (maybe a few new topics too), a hardcover with color diagrams, and hopefully a few more chuckles[1] that mostly only we nerdy types will understand.

If you want to know what this book is all about, here is my original preface:

> The intent of this book is to cover the basics that I believe have been either left out of your education or forgotten over time. Hopefully it will become one of those well-worn texts that you drop on the desk of the new guy when he asks you a question. There is something for every student, engineer, manager, and teacher in electrical engineering. My mantra is, "It ain't all that hard!" Years ago I had a counselor in college tell me proudly that they flunked out over half the students who started the engineering program. Needing to stay on her good side, I didn't say much at the time. I always wondered, though. If you fail so many students, isn't that really a failure to teach the subject well? I say "It ain't all that hard" to emphasize that even a hick with bad grammar like me can understand the world of electrical engineering. This means you can too! I take a different stance than that counselor of years ago, asserting that everyone who wants to can understand this subject. I believe that way more than 50% of the people who read this book will get something out of it. It would be nice to show the statistics to that counselor some day; she was encouraging me to drop out when she made her comment. So good luck, read on, and prove me right: It ain't all that hard!

[1] Just a hint, most of the chuckles are in the footnotes, and if you like those, check out the glossary too!

Well, that about says it all. If you do decide to give this book a chance, I want to say thank you, and I hope it brings you success in all you do!

OVERVIEW

For Engineers

Granted, there are many good teachers out there and you might have gotten the basics, but time and too many "status reports" have dulled the finish on your basic knowledge set. If you are like me, you have found a few really good books that you often pull off the shelf in a time of need. They usually have a well-written, easy-to-understand explanation of the particular topic you need to apply. I hope this will be one of those books for you.

You might also be a fish out of water, an ME thrown into the world of electrical engineering, and you would really like a basic understanding to work with the EEs around you. If you get a really good understanding of these principles, I guarantee you will surprise at least some of the "sparkies" (as I like to call them) with your intuitive insights into problems at hand.

For Students

I don't mean to knock the collegiate educational system, but it seems to me that too often we can pass a class in school with the "assimilate and regurgitate" method. You know what I mean: Go to class, soak up all the things the teacher wants you to know, take the test, say the right things at the right time, and leave the class without an ounce of applicable knowledge. I think many students are forced into this mode when teachers do not take the time to lay the groundwork for the subject they are covering. Students are so hard-pressed to simply keep up that they do not feel the light bulb go on over their heads or say, "Aha, now I get it!" The reality is, if you leave the class with a fundamental understanding of the topic and you know that topic by heart, you will be eminently more successful applying that basic knowledge than anything from the end of the syllabus for that class.

For Managers

The job of the engineering manager[2] really should have more to it than is depicted by the pointy-haired boss you see in *Dilbert* cartoons. One thing many

[2] Suggested alternate title for this book from reader Travis Hayes: *EE for Dummies and Those They Manage*. I liked it, but I figured the pointy-haired types wouldn't get it.

managers do not know about engineers is that they welcome truly insightful takes on whatever they might be working on. Please notice I said "truly insightful"; you can't just spout off some acronym you heard in the lunchroom and expect engineers to pay attention. However, if you understand these basics, I am sure there will be times when you will be able to point your engineers in the right direction. You will be happy to keep the project moving forward, and they will gain a new respect for their boss. (They might even put away their pointy-haired doll!)

For Teachers

Please don't get me wrong, I don't mean to say that all teachers are bad; in fact mostof my teachers (barring one or two) were really good instructors. However, sometimes I think the system is flawed. Given pressures from the dean to cover *X*, *Y*, and *Z* topics, sometimes the more fundamental *X* and *Y* are sacrificed just to get to topic *Z*.

I did get a chance to teach a semester at my own alma mater. Some of these chapters are directly from that class. My hope for teachers is to give you another tool that you can use to flip the switch on the "Aha" light bulbs over your students' heads.

For Everyone

At the end of each topic discussed in this book are bullet points I like to call *Thumb Rules*. They are what they seem: those "rule-of-thumb" concepts that really good engineers seem to just know. These concepts are what always led them to the right conclusions and solutions to problems. If you get bored with a section, make sure to hit the Thumb Rules anyway. There you will get the distilled core concepts that you really should know.

About the Author

Darren Coy Ashby is a self-described "techno geek with pointy hair." He considers himself a jack-of-all-trades, master of none. He figures his common sense came from his dad and his book sense from his mother. Raised on a farm and graduated from Utah State University seemingly ages ago, Darren has nearly 20 years of experience in the real world as a technician, an engineer, and a manager. He has worked in diverse areas of compliance; production; testing; and, his personal favorite, R&D.

He jumped at a chance some years back to teach a couple of semesters at his alma mater. For about two years, he wrote regularly for the online magazine Chipcenter.com. Darren is currently the director of electronics R&D at a billion-dollar consumer products company. His passions are boats, snowmobiles, motorcycles, and pretty much anything with a motor. When not at his day job, he spends most of his time with his family and a promising R&D consulting/manufacturing firm he started a couple of years ago.

Darren lives with his beautiful wife, four strapping boys, and cute little daughter next to the mountains in Richmond, Utah. You can email him with comments, complaints, and general ruminations at dashby@raddd.com.

CHAPTER 0

What Is Electricity Really?

CHICKEN VS. EGG

Which came first, the chicken or the egg? I was faced with just such a quandary when I set down to create the original edition of this book. The way that I found people got the most out of the topics was to get some basic ideas and concepts down first; however, those ideas were built on a presumption of a certain amount of knowledge. On the other hand, I realized that the knowledge that was to be presented would make more sense if you first understood these concepts—thus my chicken-vs.-egg dilemma.

Suffice it to say that I jumped ahead to explaining the chicken (the chicken being all about using electricity to our benefit). I was essentially assuming that the reader knew what an egg was (the "egg" being a grasp on what electricity is). Truth be told, it was a bit of a cheat on my part,[1] and on top of that I never expected the book to be such a runaway success. Turns out there are lots of people out there who want to know more about the magic of this ever-growing electronic world around us. So, for this new and improved edition of the book, I will digress and do my best to explain the "egg." Skip ahead if you have an idea of what it's all about,[2] or maybe stick around to see if this is an enlightening look at what electricity really is.

---

[1]Do we all make compromises in the face of impossible deadlines? Are the deadlines only impossible because of our own procrastination? Those are both very heavy-duty questions, not unlike that of the chicken-vs.-egg debate. ☺

[2]Thus the whole Chapter 0 idea; you can argue that 0 or 1 is the right number to start counting with, so pick whichever chapter you want to begin with of these two and have at it.

SO WHAT IS ELECTRICITY?

The electron—what is it? We haven't ever seen one, but we have found ways to measure a bunch of them. Meters, oscilloscopes, and all sorts of detectors tell us how electrons move and what they do. We have also found ways to make them turn motors, light up light bulbs, and power cell phones, computers, and thousands of other really cool things.

What is electricity though? Actually, that is a very good question. If you dig deep enough you can find RSPs[3] all over the world who debate this very topic. I have no desire to that join that debate (having not attained RSP status yet). So I will tell you the way I see it and think about it so that it makes sense in my head. Since I am just a hick from a small town, I hope that my explanation will make it easier for you to understand as well.

THE ATOM

We need to begin by learning about a very small particle that is referred to as an *atom*. A simple representation of one is shown in Figure 0.1.

Atoms[4] are made up of three types of particles: protons, neutrons, and electrons. Only two of these particles have a feature that we call *charge*. The proton carries a *positive charge* and the electron carries a *negative charge,* whereas the neutron carries no charge at all. The individual protons and neutrons are much more massive than the wee little electron. Although they aren't the same size, the proton and the electron do carry equal amounts of opposite charge.

Now, don't let the simple circles of my diagram lead you to believe that this is the path that electrons move in. They actually scoot around in a more energetic 3D motion that physicists refer to as a *shell*. There are many types and shapes of shells, but the specifics are beyond the scope of this text. You do need to understand that when you dump enough *energy* into an atom, you can get an electron to pop off and move fancy free. When this happens the rest of the atom has a net positive charge[5] and the electron a net negative charge.[6] Actually they have these charges when they are part of the atom. They simply

---

[3]RSP = Really Smart Person. As you will soon learn, I do hope to get an acronym or two into everyday vernacular for the common engineer. BTW, I believe that many engineers are RSPs; it seems to be a common trait among people of that profession.

[4]The atom is really, really small. We can sorta "see" an atom these days with some pretty cool instruments, but it is kinda like the way a blind person "sees" Braille by feeling it.

[5]An atom with a net charge is also known as an *ion*.

[6]Often referred to as a *free electron*.

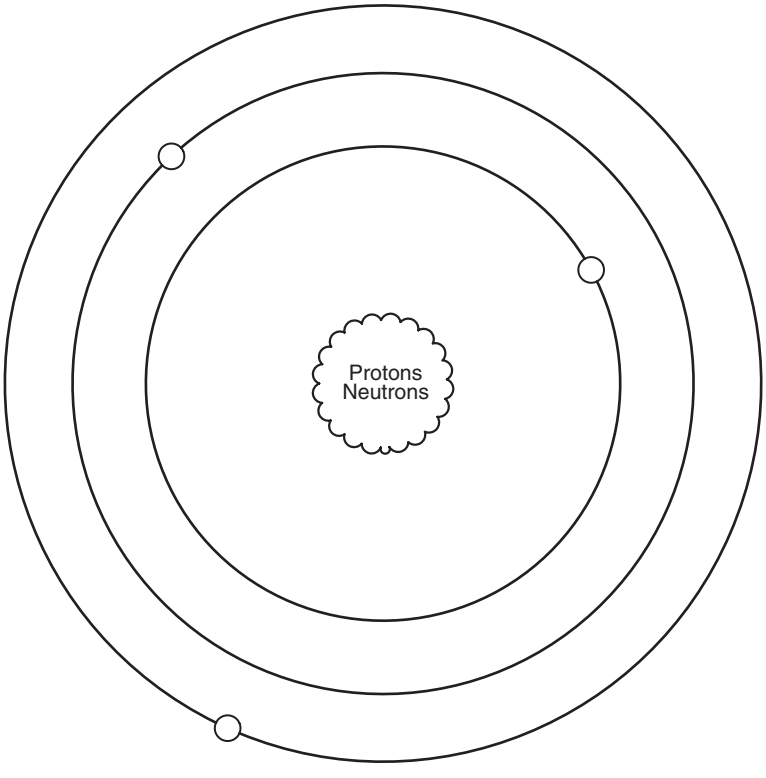


FIGURE 0.1
Very basic symbol of an atom.

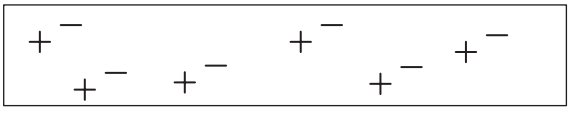


FIGURE 0.2
Electrons are "stuck" in these shells in an insulator; they can't really leave and move fancy free.

cancel each other out so that when you look at the atom as a whole the net charge is zero.

Now, atoms don't like having electrons missing from their shells, so as soon as another one comes along it will slip into the open slot in that atom's shell. The amount of energy or work it takes to pop one of these electrons loose depends on the type of atom we are dealing with. When the atom is a good insulator, such as rubber, these electrons are stuck hard in their shells. They aren't moving for anything. Take a look at the sketch in Figure 0.2.

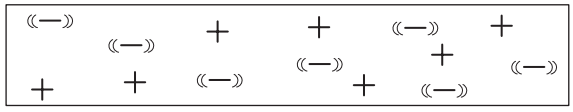


FIGURE 0.3
An electron sea.

In an insulator, these electron charges are "stuck" in place, orbiting the nucleus of the atom—kinda like water frozen in a pipe.[7] Do take note that there are just as many positive charges as there are negative charges.

With a good conductor like copper, the electrons in the outer shells of the atoms will pop off at the slightest touch; in metal elements these electrons bounce around from atom to atom so easily that we refer to them as an *electron sea*, or you might hear them referred to as *free electrons*. More visuals of this idea are shown in Figure 0.3.

You should note that there are still just as many positive charges as there are negative charges. The difference now is not the number of charges; it is the fact that they can move easily. This time they are like water in the pipe that isn't frozen but liquid—albeit a pipe that is already full of water, so to speak. Getting the electrons to move just requires a little push and away they go.[8] One effect of all these loose electrons is the silvery-shiny appearance that metals have. No wonder that the element that we call silver is one of the best conductors there is.

One more thing: A very fundamental property of charge is that like charges repel and opposite charges attract.[9] If you bring a free electron next to another free electron, it will tend to push the other electron away from it. Getting the positively charged atoms to move is much more difficult; they are stuck in place in virtually all solid materials, but the same thing applies to positive charges as well.[10]

---

[7] I like the frozen water analogy; just don't take it too far and think you just need to melt them to get them to move!

[8] Analogies are a great way to understand something, but you have to take care not to take them too far. In this case take note that you can't simply tip your wire up and get the electrons to fall out, so it isn't *exactly* like water in a pipe.

[9] It strikes me that this is somewhat fundamental to human relationships. "Good" girls are often attracted to "bad" boys, and many other analogies that come to mind.

[10] There are definitely cases where you can move positive charges around. (In fact, it often happens when you feel a shock.) It's just that most of the types of materials, circuits, and so on that we deal with in electronics are about moving the tiny, super-small, commonly easy-to-move electron. For that other cool stuff, I suggest you find a good book on electromagnetic physics.

Thumb Rules

- Electricity is fundamentally charges, both positive and negative.
- Energy is work.
- There are just as many positive as negative charges in both a conductor and an insulator.
- In a good conductor, the electrons move easily, like liquid water.
- In a good insulator, the electrons are stuck in place, like frozen water (but not exactly; they don't "melt").
- Like charges repel and opposite charges attract.

NOW WHAT?

So now we have an idea of what insulators and conductors are and how they relate to electrons and atoms. What is this information good for, and why do we care? Let's focus on these charges and see what happens when we get them to move around.

First, let's get these charges to move to a place and stay there. To do this we'll take advantage of the cool effect that these charges have on each other, which we discussed earlier. Remember, opposite charges attract, whereas the same charges repel. There is a cool, mysterious, magical field around these charges. We call it the *electrostatic field*. This is the very same field that creates everything from static cling to lightning bolts. Have you ever rubbed a balloon on your head and stuck it on the wall? If so you have seen a demonstration of an electrostatic field. If you took that a little further and waved the balloon closely over the hair on your arm, you might notice how the hairs would track the movement of the balloon. The action of rubbing the balloon caused your head to end up with a net total charge on it and the opposite charge on the balloon. The act of rubbing these materials together[11] caused some electrons to move from one surface to the other, *charging* both your head and the balloon.

This electrostatic field can exert a force on other things with charges. Think about it for a moment: If we could figure out a way to put some charges on one end of our conductor, that would push the like charges away and in so doing cause those charges to move.

[11] Fun side note: Google this balloon-rubbing experiment and see what charge is where. Also research the fact that this happens more readily with certain materials than others.

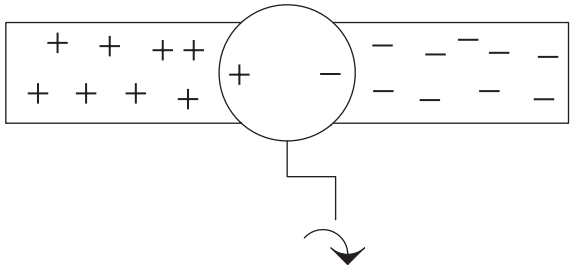


FIGURE 0.4
Hypothetical electron pump.

Figure 0.4 shows a hypothetical device that separates these charges. I will call it an *electron pump* and hook it up to our copper conductor we mentioned previously.

In our electron pump, when you turn the crank, one side gets a surplus of electrons, or a negative charge, and on the other side the atoms are missing said electrons, resulting in a positive charge.[12]

If you want to carry forward the water analogy, think of this as a pump hooked up to a pipe full of water and sealed at both ends. As you turn the pump, you build up pressure in the pipe—positive pressure on one side of the pump and negative pressure on the other. In the same way, as you turn the crank you build up charges on either side of the pump, and then these charges push out into the wire and sit there because they have no place to go. If you hook up a meter to either end you would measure a potential (think difference in charge) between the two wires. That potential is what we call *voltage*.

NOTE

It's important to realize that it is by the nature of the location of these charges that you measure a voltage. Note that I said *location*, not *movement*. Movement of these charges is what we call *current*. (More on that later.) For now what you need to take away from this discussion is that it is an accumulation of charges that we refer to as *voltage*. The more like charges you get in one location, the stronger the electrostatic field you create.[13]

[12]There is actually a device that does this. It is called a *Van de Graaff generator,* so it really isn't hypothetical, but I really like the word *hypothetical*. Just saying it seems to raise my IQ!

[13]There isn't a good water analogy for this field. You simply need to know it is there; it is important to understand that this field exists. If you still don't grasp this field, get a balloon and play with it till you do. Remember, even the best analogies can break down. The point is to use the analogy to help you begin to grasp the topic, then experiment till you understand all the details.

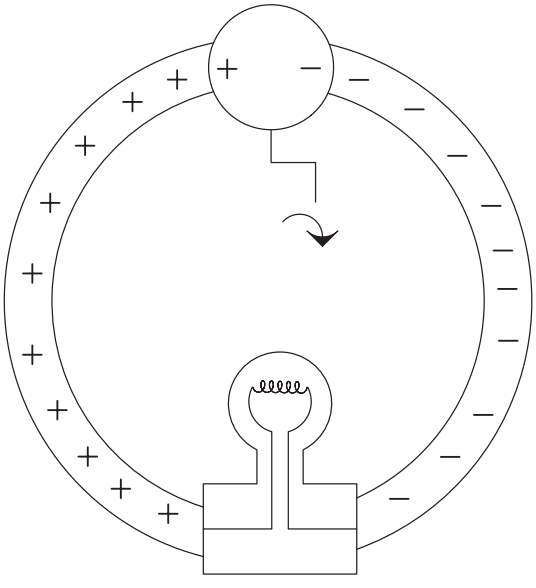


FIGURE 0.5
Electron pump with light bulb.

Okay, it's later now. We find that another very cool thing happens when we move these charges. Let's go back to our pump and stick a light bulb on the ends of our wires, as shown in Figure 0.5.

Remember that opposite charges attract? When you hook up the bulb, on one side you have positive charges, on the other negative. These charges push through the light bulb, and as they do they heat up the filament and make it light up. If you stop turning the electron pump, this potential across the light bulb disappears and the charges stop moving. Start turning the pump and they start moving again. The movement of these charges is called *current*.[14] The really cool thing that happens is that we get another invisible field that is created when these charges move; it is called the *electromagnetic field*. If you have ever played with a magnet and some iron filings, you have seen the effects of this field.[15]

So, to recap, if we have a bunch of charges hanging out, we call it *voltage*, and when we keep these charges in motion we call that *current*. Some typical water analogies look at voltage as pressure and current as flow. These are helpful to

[14]Current is coulombs per sec, a measure of flow that has units of amperes, or amps.
[15]In a permanent magnet, all the electrons in the material are scooting around their respective atoms in the same direction; it is the movement of these charges that creates the magnetic field.

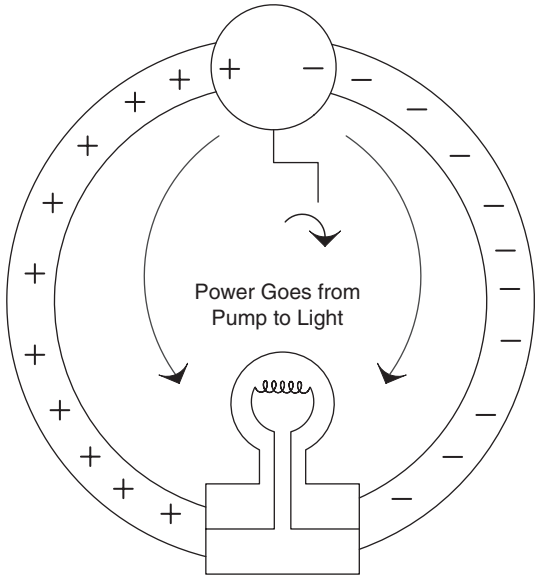


FIGURE 0.6
The electromagnetic and electronic fields transmit the work from the crank to the light bulb.

grasp the concept, but keep in mind that a key thing with these charges and their movements is the seemingly magical fields they produce. Voltage generates an electrostatic field (it is this field repelling or attracting other charges that creates the voltage "pressure" in the conductor). Current or flow or movement of the charges generates a magnetic field around the conductor. It is very important to grasp these concepts to enhance your understanding of what is going on. When you get down to it, it is these fields that actually move the work or energy from one end of a circuit to another.

Let's go back to our pump and light bulb for a minute, as shown in Figure 0.6.

Turn the pump and the bulb lights up. Stop turning and it goes out. Start turning and it immediately lights up again. This happens even if the wires are long! We see the effect immediately. Think of the circuit as a pair of pulleys and a belt. The charges are moving around the circuit, transferring power from one location to another—see Figure 0.7.[16]

Fundamentally, we can think of the concept as shown in the drawing in Figure 0.8.

[16]This diagram is a simplified version of a *scalar wave* diagram. I won't go into scalar diagrams in depth here, to limit the amount of information you need to absorb. However, I do recommend that you learn about these when you feel ready.

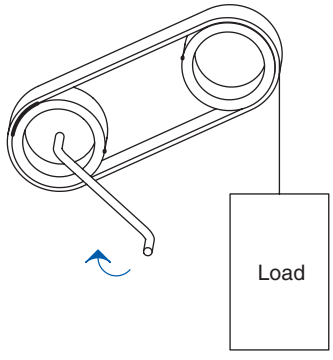


FIGURE 0.7
The belt transmits the work from the crank to the load.

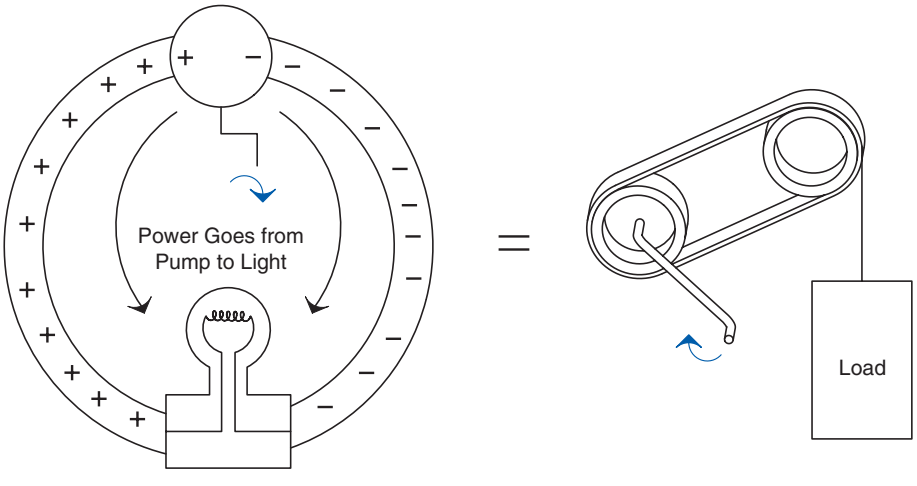


FIGURE 0.8
The cool magical fields act like the belt transmitting what we call energy, work, or power.

Even if the movement of the belt is slow,[17] we see the effects on the pulley immediately, at the moment the crank is turned. It is the same way with the light bulb. However, the belt is replaced by the circuit, and it is actually the

[17]In fact the charges in the wire are moving much more slowly than one might think. In fact, DC current moves at about 8 CM per hour. (In a typical wire, exact speed depends on several factors, but it is much slower than you might think.) AC doesn't even keep flowing, it just kinda bounces back and forth. If you think about it, you might wonder how flipping a switch can get a light to turn on so quickly. Thus the motor and belt analogy; it is the fact that the wire "pipe" is filled (in the same way the belt is connected to the pulley) with these charges that creates the instantaneous effect of a light turning on.

electromagnetic[18] fields pushing charges around that transmit the work to the bulb. Without the effects of both these fields, we couldn't move the energy input at the crank to be output at the light bulb. It just wouldn't happen.

Like the belt on the pulleys, the charges move around in a loop. But the work that is being done at the crank moves out to the light bulb, where it is used up making the light shine. Charges weren't used up; current wasn't used up. They all make the loop (just like the belt in the pulley example). It is *energy* that is used up. Energy is work; you turning the crank is work. The light bulb takes energy to shine. In the bulb energy is converted into heat on the filament that makes it glow so bright that you get light. But remember, it is energy that it takes to make this happen. You need both voltage and current (along with their associated fields) to transfer energy from one point to another in an electric circuit.

Thumb Rules

- An accumulation of charges is what we call *voltage.*
- Movement of charges is what we call *current* or *amperage.*
- Energy is work; in a circuit the electromagnetic effects move energy from one point to another.

A PREVIEW OF THINGS TO COME

Now, all the electronic items that we are going to learn about are based on these charges and their movement. We will learn about *resistance*—the measurement of how difficult it is to get these electrons to pop loose and move around a circuit. We will learn about a *diode*, a device that can block these charges from moving in one direction while letting them pass in another. We will learn about a *transistor* and how (using principles similar to the diode) it can switch a current flow on and off.[19]

---

[18] When I use the term *electromagnetic*, it is referring to the effects of both the electrostatic field and the magnetic field that we have been talking about.

[19] These are called *semiconductors*, and with good reason: They lie somewhere (*semi-*) between an insulator and a conductor in their ability to move charges. As you will learn later, we capitalize on this fact and the cool effects that occur when you jam a couple of different types together.

We will learn about generators and batteries and find out they are simply different versions of the *electron pump* that we just talked about.

We will learn about motors, resistors, lights, and displays—all items that consume the power that comes from our electron pump.

But just remember, it all comes back to this basic concept of a charge, the fields around it when it sits there, and the fields that are created when the charges *move*.

IT JUST SEEMS MAGICAL

Once you grasp the idea of charges and how the presence and movement of these charges transfer energy, the magic of electricity is somewhat lost. If you get the way these charges are similar to a belt turning a pulley, you are already further ahead in understanding than I was when I graduated from college. Whatever you do, don't let anyone tell you that you can't learn[20] this stuff. It really isn't all that magical, but it does require you to have an imagination. You might not be able to see it, but you surely can grasp the fundamentals of how it works.

So give it a try; don't say you can't do this,[21] because I am sure you can. If you read this book and don't come away with a better grasp of all things electrical and electronic, please drop me a line and complain about it. As long as my inbox isn't too clogged by email from all those raving reviews, I will be sure to get back to you.

Thumb Rules

- "Can't" is a sucker too lazy to try.
- Laziness is the mother of invention.

[20] Am I alone in my distaste for so-called weed-out courses? You know, the ones that they put in the curriculum to get people to quit because they make them so hard. I personally believe that the goal of teachers should be to teach. It follows that the goal of a university should be to teach better, not just turn people away.

[21] My dad always said, "Can't is a sucker to lazy to try!" after learning this, I also went on to develop a personal belief that laziness is the mother of invention. Does that mean the most successful inventors are those that are lazy enough to look for an easier way, but not so lazy as to try it?

CHAPTER 1

Three Things They Should Have Taught in Engineering 101

Do you remember your engineering introductory course? At most, I'll venture that you are not sure you even had a 101 course. It's likely that you did and, like the course I had, it really didn't amount to much. In fact, I don't remember anything except that it was supposed to be an "introduction to engineering."

Much later in my senior year and shortly after I graduated, I learned some very useful general engineering methodologies. They are so beneficial that I sincerely wish they had taught these three things from the beginning of my coursework. In fact, it is my belief that this should be basic, *basic* knowledge that any aspiring engineer should know. I promise that by using these in your day-to-day challenges you will be more successful and, besides that, everyone you work with will think you are a genius. If you are a student reading this, you will be amazed at how many problems you can solve with these skills. They are the fundamental building blocks for what is to come.

UNITS COUNT!

This is a skill that one of my favorite teachers drilled into me during my senior year. Till I understood unit math, I forced myself to memorize hundreds of equations just to pass tests. After applying this skill I found that, with just a few equations and a little algebra, you could solve nearly any problem. This was definitely an "Aha" moment for me. Suddenly the world made sense. Remember those dreaded story problems that you had to do in physics? Using

unit math, those problems become a breeze; you can do them without even breaking a sweat.

Unit Math

With this process the units that the quantities are in become very important. You don't just toss them aside because you can't put them in your calculator. In fact, you figure out the units you want in your answer and then work the problem backward to figure out what you need to solve it. You do all this before you do anything with the numbers at all. This basic concept was taught way back in algebra class, but no one told you to do it with units. Let's look at a very simple example.

You need to know how fast your car is moving in miles per hour (mph). You know it traveled one mile in one minute. The first thing you need to do is figure out the units of the answer. In this case it is mph, or miles per hour. Now write that down (remember *per* means divided by).

$$answer = something \cdot \frac{miles}{hour}$$

Now arrange the data that you have in a format that will give you the units you want in the answer:

$$1 \cdot mile \times \frac{1}{1 \cdot \text{min}} \times \frac{60 \cdot \text{min}}{1 \cdot hour} = answer$$

Remember, whatever is above the dividing line cancels out whatever is the same below the line, something like this:

$$1 \cdot mile \times \frac{1}{1 \cdot \cancel{\text{min}}} \times \frac{60 \cdot \cancel{\text{min}}}{1 \cdot hour} = answer$$

When all the units that can be removed are gone, what you are left with is 60 mph, which is the correct answer. Now, you might be saying to yourself that that was easy. You are right! That is the point after all—we want to make it easier. If you follow this basic format, most of the "story problems" you encounter every day will bow effortlessly to your machinations.

Another excellent place to use this technique is for solution verification. If the answer doesn't come out in the right units, most likely something was wrong in your calculation. I always put units on the numbers and equations I use in MathCad (a tool no engineer should be without). To see the correct units when all is said and done it confirms that the equations are set up properly. (The nice

thing is that MathCad automatically handles the conversions that are often needed.) So, whenever you come upon a question that seems to have a whole pile of data and you have no idea where to begin, first figure out which units you want the answer in. Then shape that pile of data till the units match the units needed for the answer.

REMEMBER THIS

By letting the units mean something in the problem, the answer you get will actually mean something, too.

Sometimes Almost Is Good Enough

My father had a saying: "'Almost' only counts in horseshoes and hand grenades!" He usually said this right after I "almost" put his tools away or I "almost" finished cleaning my room. Early in life I became somewhat of an expert in the field of "almost." As my dad pointed out, there are many times when almost doesn't count. However, as this bit of wisdom states, it probably is good enough to *almost* hit your target with a hand grenade. There are a few other times when almost is good enough, too. One of them is when you are trying to estimate a result. A skill that goes hand in hand with the idea of unit math is that of estimation.

The skill or art of estimation involves two main points. The first is rounding to an easy number and the second is understanding ratios and percentages. The rounding part comes easy. Let's say you are adding two numbers, 97 and 97. These are both nearly 100, so say they are 100 for a minute; add them together and you get 200, or nearly so. Now, this is a very simplified explanation of this idea, and you might think, "Why didn't you just type 97 into your calculator a couple of times and press the equals sign?" The reason is, as the problems become more and more complex, it becomes easier to make a mistake that can cause you to be far off in your analysis. Let's apply this idea to our previous example. If your calculator says 487 after you add 97 to 97, and you compare that with the estimate of 200 that you did in your head, you quickly realize that you must have hit a wrong button.

Ratios and percentages help you get an idea of how much one thing affects another. Say you have two systems that add their outputs together. In your design, one system outputs 100 times more than the other. The ratio of one to the other is 100:1. If the output of this product is way off, which of these two systems do you think is most likely at fault? It becomes obvious that one system has a bigger effect when you estimate the ratio of one to the other.

Developing the skill of estimation will help you eliminate hunting dead ends and chasing your tail when it comes to engineering analysis and troubleshooting. It will also keep you from making dumb mistakes on those pesky finals in school! Learn to estimate in your head as much as possible. It is okay to use calculators and other tools—just keep a running estimation in your head to check your work.

When you are estimating, you are trying to simplify the process of getting to the answer by allowing a margin of error to creep in. The estimated answer you get will be "almost" right, and close enough to help you figure out where else you may have screwed up.

In the game of horseshoes you get a few points for "almost" getting a ringer, but I doubt your boss will be happy with a circuit that "almost" works. However, if your estimates are "almost" right, they can help you design a circuit that even my dad would think is good enough.

Thumb Rules

- Always consider units in your equations; they can help you make sure you are getting the right answer.
- Use units to create the right equation to solve the problem. Do this by making a unit equation and canceling units until you have the result you want.
- Use estimation to determine approximately what the answer should be as you are analyzing and troubleshooting; then compare that to the results to identify mistakes.

HOW TO VISUALIZE ELECTRICAL COMPONENTS

Mechanical engineers have it easy. They can see what they are working on most of the time. As an EE, you do not usually have that luxury. You have to imagine how those pesky electrons are flittering around in your circuit. We are going to cover some basic comparisons that use things you are familiar with to create an intuitive understanding of a circuit. As a side benefit, you will be able to hold your own in a mechanical discussion as well. There are several reasons to do this:

- The typical person understands the physical world more intuitively than he understands the electrical one. This is because we interact with the

physical world using all our senses, whereas the electrical world is still very magical, even to an educated engineer. This is because much of what happens inside a circuit cannot be seen, felt, or heard. Think about it. You flip on a light switch and the light goes on. You really don't consider how the electricity caused it to happen. Drag a heavy box across the floor, and you certainly understand the principle of friction.

- The rules for both disciplines are exactly the same. Once you understand one, you will understand the other. This is great, because you only have to learn the principles once. In the world of Darren we call EEs "sparkies" and MEs "wrenches." If you grok[1] this lesson, a "sparky" can hold his own with the best "wrench" around, and vice versa.
- When you get a feel for what is happening inside a circuit, you can be an amazingly accurate troubleshooter. The human mind is an incredible instrument for simulation, and unlike a computer, it can make intuitive leaps to correct conclusions based on incomplete information. I believe that by learning these similarities you increase your mind's ability to put together clues to the operation and results of a given system, resulting in correct analysis. This will help your mind to "simulate" a circuit.

Physical Equivalents of Electrical Components

Before we move on to the physical equivalents, let's understand voltage, current, and power. *Voltage* is the potential of the charges in the circuit. *Current* is the amount of charge flowing[2] in the circuit. Sometimes the best analogies are the old overused ones, and that is true in this case. Think of it in terms of water in a squirt gun. Voltage is the amount of pressure in the gun. Pressure determines how far the water squirts, but a little pea shooter with a 30-foot shot and a dinky little stream won't get you soaked. Current is the size of the water stream from the gun, but a large stream that doesn't shoot far is not much help in a water fight. What you need is a super-soaker 29 gazillion, with a half-inch water stream that shoots 30 feet. Now that would be a *powerful* water-drenching weapon. Voltage, current, and power in electrical terms are related the same way. It is in fact a simple relationship; here is the equation:

$$voltage * current = power \qquad \textbf{Eq. 1.1}$$

[1] *Grok* means to understand at a deep and personal level. I highly suggest reading Robert Heinlein's *Stranger in a Strange Land* for a deeper understanding of the word *grok*.
[2] Or moving.

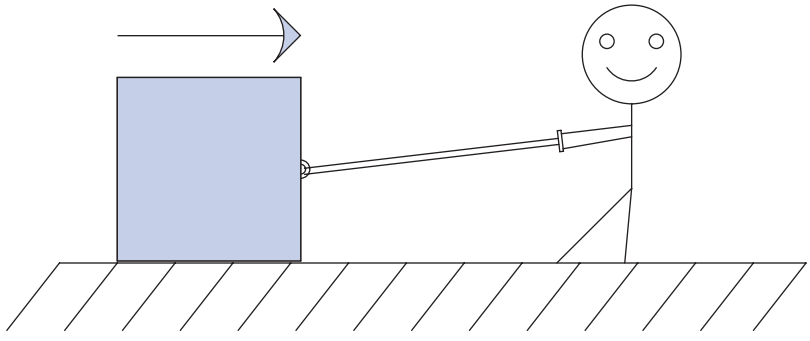


FIGURE 1.1
Friction resists smiley stick boy's efforts.

To get power, you need both voltage and current. If either one of these is zero, you get zero power output. Remember, power is a combination of these two items: current and voltage.

Now let's discuss three basic components and look at how they relate to voltage and current.

The Resistor Is Analogous to Friction

Think about what happens when you drag a heavy box across the floor, as shown in Figure 1.1. A force called *friction* resists the movement of the box. This friction is related to the speed of the box. The faster you try to move the box, the more the friction resists the movement. It can be described by an equation:

$$friction = \frac{force}{speed} \qquad \textbf{Eq. 1.2}$$

Furthermore, the friction dissipates the energy loss in the system with heat. Let me rephrase that. Friction makes things get warm. Don't believe me? Try rubbing your hands together right now. Did you feel the heat? That is caused by friction. The function of a resistor in an electrical circuit is equal to friction. The resistor resists the flow of electricity* just like friction resists the speed of the box. And, guess what? It heats up as it does so. An equation called Ohm's Law describes this relationship:

$$resistance = \frac{voltage}{current} \qquad \textbf{Eq. 1.3}$$

*Resistance represents the amount of effort it takes to pop one of those pesky electrons we talked about in chapter 0 and to move it to the atom next to it.

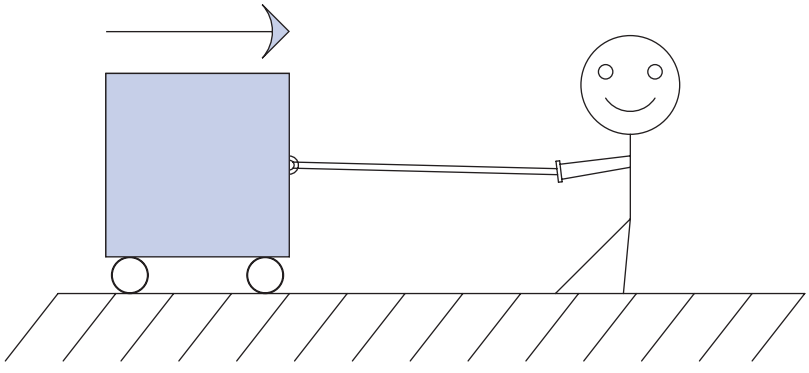


FIGURE 1.2
Wheels eliminate friction, but smiley has a hard time getting it up to speed and stopping it.

Do you see the similarity to the friction equation? They are exactly the same. The only real difference is the units you are working in.

The Inductor Is Analogous to Mass

Let's stay with the box example for now. First let's eliminate friction, so as not to cloud our comprehension. The box shown in Figure 1.2 is on a smooth track with virtually frictionless wheels. You notice that it takes some work to get the box going, but once it's moving, it coasts along nicely. In fact, it takes work to get it to stop again. How much work, depends on how heavy the box is. This is known as the *law of inertia*. Newton postulated this idea long before electricity was discovered, but it applies very well to inductance. Mass resists a *change* in speed. Correspondingly, inductance resists a *change* in current.

$$mass = \frac{force * time}{speed} \qquad \textbf{Eq. 1.4}$$

$$inductance = \frac{voltage * time}{current} \qquad \textbf{Eq. 1.5}$$

The Capacitor Is Analogous to a Spring

So what does a spring do? Take hold of a spring in your mind's eye. Stretch it out and hold it, and then let it go. What happens? It snaps back into position, as shown in Figure 1.3 on the next page. A spring has the capacity to store energy. When a force is applied, it will hold that energy till it is released. *Capacitance* is similar to the elasticity of the spring. (One note: The spring constant that

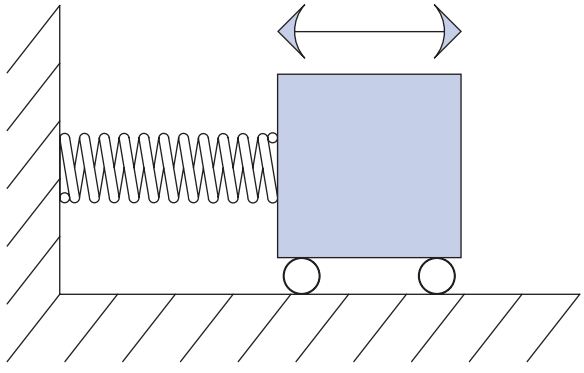


FIGURE 1.3
Get this started and it will keep bouncing until friction brings it to a halt.

you might remember from physics texts is the inverse of the elasticity.) I always thought it was nice that the word *capacitor* is used to represent a component that has the *capacity* to store energy.[3]

$$spring = \frac{speed * time}{force} \qquad \text{Eq. 1.6}$$

$$capacitance = \frac{current * time}{force} \qquad \text{Eq. 1.7}$$

A Tank Circuit

Take the basic tank or LC circuit. What does it do? It oscillates. A perfect circuit would go on forever at the resonant frequency. How should this appear in our mechanical circuit? Think about the equivalents: an inductor and a capacitor, a spring and mass. In a thought experiment, hook the spring up to the box from the previous drawing. Now give it a tug. What happens? It oscillates.

A Complex Circuit

Let's follow this reasoning for an LCR circuit. All we need to do is add a little resistance, or friction, to the mass-spring of the tank circuit. Let's tighten the wheels on our box a little too much so that they rub. What will happen after

[3]Technically, an *inductor* can store energy too. In a capacitor the energy is stored in the electric field that is generated in and around the cap; in an inductor energy is stored in the magnetic field that is generated around the coils. This energy stored in an inductor can be tapped very efficiently at high currents. That is why most switching power supplies have an inductor in them as the primary passive component.

you give the box a tug? It will bounce back and forth a bit till it comes to a stop. The friction in the wheels slows it down. This friction component is called a *damper* because it dampens the oscillation. What is it that a resistor does to an LC circuit? It dampens the oscillation.

There you have it—the world of electricity reduced to everyday items. Since these components are so similar, all the math tricks you might have learned apply as well to one system as they do to the other. Remember Fourier's theorems? They were discovered for mechanical systems long before anyone realized that they work for electrical circuits as well. Remember all that higher math you used to know or are just now learning about—Laplace transforms, integrals, derivatives, etc.? It all works the same in both worlds. You can solve a mechanical system using Laplace methods just the same as an electrical circuit.

Back in the 1950s and 1960s, the government spent mounds of dough using electrical circuits to model physical systems as described before. Why? You can get into all sorts of integrals, derivatives, and other ugly math when modeling real-world systems. All that can get jumbled quickly after a couple of orders of complexity. Think about an artillery shell fired from a tank. How do you predict where it will land? You have the friction of the air, the mass of the shell, the spring of the recoil. Instead of trying to calculate all that math by hand, you can build a circuit with all the various electrical components representing the mechanical ones, hook up an oscilloscope, and fire away. If you want to test 1000 different weights of artillery at different altitudes, electrons are much cheaper than gunpowder.[4]

Thumb Rules

- It takes voltage and current to make power.
- A resistor is like friction: It creates heat from current flow (resisting it), proportional to voltage measured across it.
- An inductor is like a mass.
- A capacitor is like a spring.
- The inductor is the inverse of the capacitor.

[4]Of course, you still had to swap out the components for the various values you were looking for. I suppose that is one reason the reign of the analog computer was so short. Once reduced to equations and represented digitally, the simulations could be varied at the click of a mouse; we just needed the digital bandwidth to increase far enough to make it feasible.

LEARN AN INTUITIVE APPROACH

Intuitive Signal Analysis

I'm not sure if *intuitive signal analysis* is actually taught in school; this is my name for it. It is something I learned on my own in college and the workplace. I didn't call it an actual discipline until I had been working for a while and had explained my methods to fellow engineers to help them solve their own dilemmas. I do think, however, that a lot of so-called bright people out there use this skill without really knowing it or putting a name to it. They seem to be able to point to something you have been working on for hours and say, "Your problem is there." They just seem to intuitively know what should happen. I believe that this is a skill that can and should be taught.

There are three underlying principles needed to apply intuitive signal analysis. (Let's just call it *ISA*. After all, if I have any hope of this catching on in the engineering world, it has to have an acronym!)

1. *You must drill the basics.* For example, what happens to the impedance of a capacitor as frequency increases? It goes down. You should know that type of information off the top of your head. If you do, you can identify a high-pass or low-pass filter immediately. How about the impedance of an inductor—what does it do as frequency increases? What does negative feedback do to an op-amp; how does its output change? You do not necessarily need to know every equation by heart, but you do need to know the direction of the change. As far as the magnitude of the change is concerned, if you have a general idea of the strength of the signal, that is usually enough to zero in on the part of the circuit that is not doing what you want it to.

2. *You need experience and lots of it.* You need to get a feel for how different components work. You need to spend a lot of time in the lab, and you need to understand the basics of each component. You need to know what a given signal will do as it passes through a given component. Remember the physical equivalents of the basic components? These are the building blocks of your ability to visualize the operation of a circuit. You must imagine what is happening inside the circuit as the input changes. If you can visualize that, you can predict what the outputs will do.

3. *Break the problem down.* "How do you eat an elephant?" the knowledge seeker asked the wise old man. "One bite at a time," old man replied. Pick a point to start and walk though it. Take the circuit and

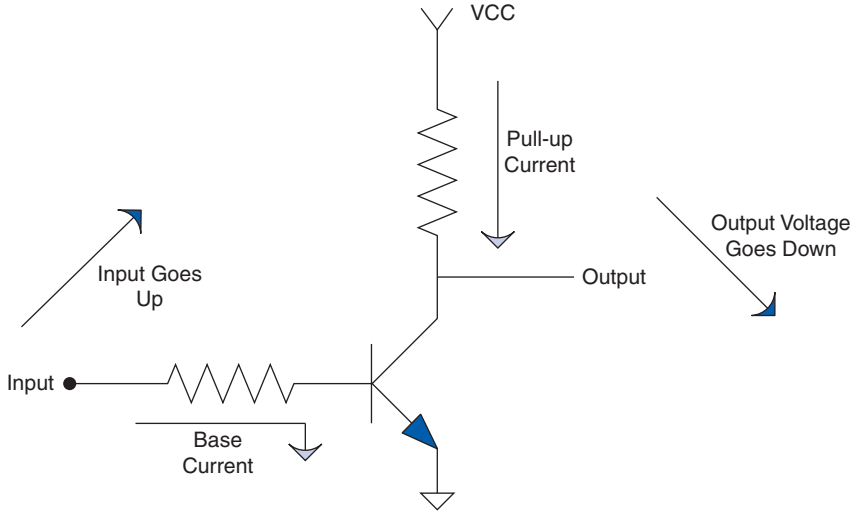


FIGURE 1.4
Use arrows to visualize what is happening to voltage and current.

break it down into smaller chunks that can be handled easily. Step by step draw arrows that show the changes of signals in the circuit, as shown in Figure 1.4. "Does current go up here?" "Voltage at such and such point should be going down." These are the types of questions and answers you should be mumbling to yourself.[5] Again, one thing you do not need to know is what the output will be *precisely*. You do not need to memorize every equation in this book to intuitively know your circuit, but you do need to know what effect changing a value of a component will have. For example, given a low-pass RC filter and an AC signal input, if you increase the value of the capacitor, what should happen to the amplitude of the output? Will it get smaller or larger?

You should know immediately with something this basic that the answer is "smaller." You should also know that how much smaller depends on the frequency of the signal and the time constant of the filter. What happens as you increase current into the base of a transistor? Current through the collector increases. What happens to voltage across a resistor as current decreases? These are simple effects of components, but you would be surprised at how many engineers don't know the answers to these types of questions off the top of their heads.

[5]Based on extensive research of talking to two or three people, I have concluded that all intelligent people talk to themselves. Whether or not they are considered socially acceptable depends on the audibility of this voice to others around them.

Spending a lot of time in the lab will help immensely in developing this skill. If you look at the response of a lot of different circuits many, many times, you will learn how they should act. When this knowledge is integrated, a wonderful thing happens: Your head becomes a circuit simulator. You will be able to sum up the effects caused by the various components in the circuit and intuitively understand what is happening. Let me show you an example.

Now, at this time you might not have a clue as to what a transistor is, so you might need to file this example away until you get past the transistor chapter, but be sure to come back to it so that the "Aha!" light bulb clicks on over your head. The analysis idea is what I am trying to get across; you need it early on, but it creates a type of chicken-and-egg dilemma when it comes to an example. So, for now consider this example with the knowledge that the transistor is a device that moves current through the output that is proportional to the current through the base.

As voltage at the input increases, base current increases. This causes the pull-up current in the resistor to increase, resulting in a larger voltage drop across the pull-up resistor. This means the voltage at the output must *go down* as the voltage at the input goes up. That is an example of putting it all together to really understand how a circuit works.

One way to develop this intuitive understanding is by using computer simulators. It is easy to change a value and see what effect it has on the output, and you can try several different configurations in a short amount of time. However, you have to be careful with these tools. It is easy to fall into a common trap: trusting the simulator so much that you will think there is something wrong with the real world when it doesn't work right in the lab. The real world is not at fault! It is the simulator that is missing something. I think it is best for the engineer to begin using simulators to model simple circuits. Don't jump into a complex model until you grasp what the basic components do—for example, modeling a step input into an RC circuit. With a simple model like this, change the values of R and C to see what happens. This is one way an engineer can develop the correct intuitive understanding of these two components. One word of warning, though: Don't spend all your time on the simulator. Make sure you get some good bench time, too.

You will find this signal analysis skill very useful in diagnosing problems as well as in your design efforts. As your intuitive understanding increases, you will be able to leap to correct conclusions without all the necessary facts. You will know when you are modeling something incorrectly, because the result

just won't look right. Intuition is a skill no computer has, so make sure you take advantage of it!

Thumb Rules

- Drill the basics; know the basic formulas by heart.
- Get a lot of experience with basic circuits; the goal is to intuitively know how a signal will be affected by a component.
- Break the problem down; draw arrows and notes on the schematic that indicate what the signal is doing.
- Determine in which direction the signal is going; is it inversely related or directly related?
- Develop estimation abilities.
- Spend time on the bench with a scope and simple components.

"LEGO" ENGINEERING

Building Blocks

Okay, so I came up with a fourth item.[6] One of my engineering instructors (we'll call him Chuck[7]) taught me a secret that I would like to pass on. Almost every discipline is easier to understand than you might think. The secret professors don't want you to know is that there are usually about five or six basic principles or equations that lie at the bottom of the pile, so to speak. These fundamentals, once they are grasped, will allow you to derive the rest of the principles or equations in that field. They are like the old simple Legos®; you had five or six shapes to make everything. If you truly understand these few basic fundamentals in a given discipline, you will excel in that discipline. One other thing Chuck often said was that all the great discoveries were only one or two levels above these fundamentals. This means that if you really know the basics well, you will excel at the rest. One thing you can be sure of is the

[6]For those of you who have been wondering if I can count.

[7]"Dr. Charles Tinney" was what he wrote on the chalkboard the first day of class. Then he turned around and said, "You can call me Chuck!" I have to credit Dr. Tinney; he was the best teacher I have ever had. For him nothing was impossible to understand or to teach you to understand.

human tendency to forget. All the higher-level stuff is often left unused and will quickly be forgotten, but even an engineer-turned-manager like me uses the basics nearly every day.

Since this is a book on electrical engineering, let's list the fundamental equations for electrical circuits as I see them:

- Ohm's Law
- Voltage divider rule
- Capacitors impede changes in voltage
- Inductors impede changes in current
- Series and parallel resistors
- Thevenin's theorem

We will get into these concepts in more detail later in the chapters, but let me touch on a couple of examples. You might say, "You didn't even list series and parallel capacitors. Isn't that a basic rule?" Well, you are right, it is fairly basic, but it really isn't at the bottom of the pile. Series and parallel resistors are even more fundamental because all that really happens when you add in the caps is that the frequency of the signal is taken into account; other than that it is exactly the same equation! You would be better served to understand how a capacitor or inductor works and apply it to the basics than to try to memorize too many equations. "What about Norton's theorem?" you might ask. Bottom line, it is just the flip side of Thevenin's theorem, so why learn two when one will do? I prefer to think of it in terms of voltage, so I set this to memory. You could work in terms of current and use Norton's theorem, but you would arrive at the same answer at the end of the day. So pick one and go with it.

You can always look up the more advanced stuff, but most of the time a solid application of the basics will force the problem at hand to submit to your engineering prowess. These six rules are things that you should memorize, understand, and be able to do approximations of in your head. These are the rules that will make the intuition you are developing a powerful tool. They will unleash the simulation capability that you have right in your own brain.

If you really take this advice to heart, years down the road when you've been given your "pointy hairs"[8] and you have forgotten all the advanced stuff you used to know, you will still be able to solve engineering problems to the amazement of your engineers.

[8] In case you have lived under a rock for the last few years and missed a certain very successful engineering cartoon, this means "promoted to management."

This can be generalized to all disciplines. Look at what you are trying to learn, figure out the few basic points being made, from which you can derive the rest, and you will have discovered the basic "Legos" for that subject. Those are the things you should know forward and backward to succeed in that field. Besides, Legos are fun, aren't they?

Thumb Rules

- There are a few rules in any discipline from which you can derive the rest.
- Learn these rules by heart; gain an intuitive understanding of them.
- Most significant discoveries are only a level or two above these basics.

CHAPTER 2

Basic Theory

Every discipline has fundamentals that are used to extrapolate all the other, more complex ideas. Basics are the most important thing you can know. It is knowledge of the basics that helps you apply all that stuff in your head correctly. It doesn't matter if you can handle quadratic equations and calculus in your sleep. If you don't grasp the basics, you will find yourself constantly chasing a problem in circles without resolution. If you get anything out of this text, make sure that you really understand the basics!

OHM'S LAW STILL WORKS: CONSTANTLY DRILL THE FUNDAMENTALS

Ohm's Law

This, I believe, is one of the best-taught principles in school for the budding engineer or technician, and it should be. So why go over it? Well, two reasons come to mind: One, you can't go over the basics too much, and two, though any engineer can quote Ohm's Law by heart, I have often seen it ignored in application.

First, let's state Ohm's Law: Voltage equals current multiplied by resistance; it is shown in Figure 2.1 on the next page.

It is simple, but do you consider that resistance exists in every part of a circuit[1]? It is easy to forget that, especially since many simulators do. I think

[1]Okay, you could be all snitty here and point out that superconductors by definition don't have resistance. But then my cool story coming up wouldn't have the impact needed to drive home this point that applies to 99.999999% of all circuits out there!

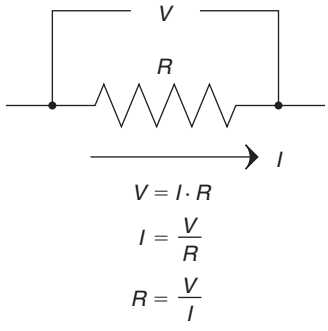


FIGURE 2.1
Ohm's Law, the heart of all things electrical.

the best way to drive this point home is to recount the way it was driven home to me.

There I was—a lowly engineering student. I was working as a technician or associate engineer (depending on whom you asked). I was arguing with my boss, who had an MSEE degree, but he just wouldn't believe me; neither would my lead engineer (who had a BSEE). I couldn't bring myself to distrust Ohm's Law, even in light of their "superior" knowledge. I'd had less heated debates with rabid dogs. This was the problem: Our department needed to measure the current of a DC motor that could range from 5 A to 15 A at any given time, but our multimeters had a 10 A fuse in the current measuring circuit.

So, using Ohm's Law (which was fresh in my mind, being a student and all), I designed a shunt to measure current. I wanted to get a good reading but disturb the circuit as little as possible, so I chose a 0.1 Ω resistor. I built a box to house it and installed banana-jack plugs to provide an easy interface to a voltmeter. The design looked like the one shown in Figure 2.2.

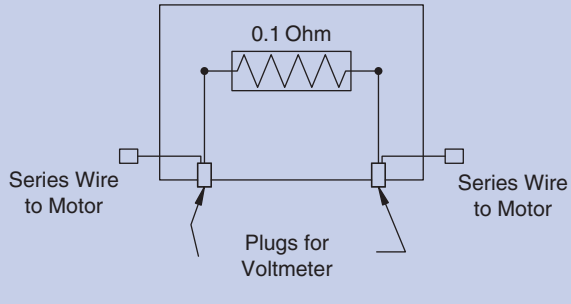


FIGURE 2.2
Original design of simple current-measuring circuit.

Everyone thought it was a great idea, so I built a couple of boxes and we started using them right away. After a while, however, we noticed that they were not very accurate. Sometimes they would be off by as much as 50 to 60%. No one could figure out why, so I sat down to analyze what I had created.

After a few minutes, I said to myself, "Well, duhhh!" I realized that to make the assembly easy I had soldered the wires from the motor to the banana jacks and then soldered some short 14-gauge jumpers to the shunt resistor. My circuit really looked like the drawing shown in Figure 2.3.

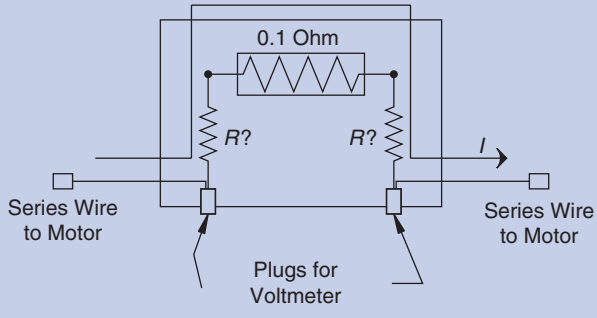


FIGURE 2.3
As-built simple current-measuring circuit.

My voltmeter was measuring across a larger resistance value than 0.1 ohms. Wire has resistance, too; even a couple of inches of 14-gauge wire has a few hundredths of an ohm. Remembering Ohm's Law:

$$V = I * R \qquad \text{Eq. 2.1}$$

I realized that this means if you increase R, you get more V for the same amount of current, leading to the errors we were seeing. I had made a simple mistake that fortunately was easy to correct. I redesigned the box on paper to look like the drawing in Figure 2.4.

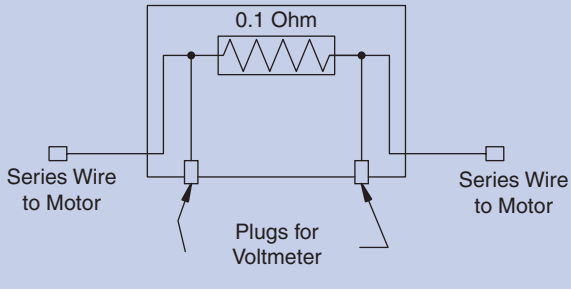


FIGURE 2.4
Redesigned current-measuring circuit.

(*Continued*)

I took this to my boss (the one with the MSEE who could do math in his head that I would only attempt with MathCad and a cold drink). His reaction floored me. He reviewed it with the lead engineer and they came to the conclusion that I was completely wrong. They were talking about things like temperature coefficients and phase shifts in current and RMS and a bunch of other topics that were over my head at the time. Thus began the argument. I explained that two points on a schematic had to be connected by a wire and a wire had resistance. Though it is often ignored, it was significant in this case because the shunt resistor was such a small value.

As they hemmed and hawed over this, I learned that many times it is human nature to ignore what one learned long ago and try to apply more advanced theories just because you know them. Also, all the knowledge in the world isn't worth jack if it is incorrectly applied. I continued to press my point. I must have written Ohm's Law on the white board 50 times by then.

They finally conceded and agreed that the extra wire between the banana jack and the shunt was the cause of the error. That was not the end of the disagreement, though. How in the world was my new design going to fix the problem by simply repositioning the wires? The resistance was still in the circuit, was it not? I wrote down Ohm's Law another 100 times and explained that the current through the meter was very small, making the resistance in the wire insignificant again. My astonishment reached new levels as I observed the human ability to overlook the obvious. The first argument was nothing compared to this one. The fireworks started to really fly then.

What is the moral of this story? Well, Scott Adams, creator of *Dilbert,* said, "Everyone has moments of stupidity," as he watched someone fix his "broken" pager by putting in a new battery. I have to agree with him. I rediscover Ohm's Law about every 6 months. Always, always, always check the basics before you start looking for more complicated solutions! My father, a mechanic, tells a story of rewiring an entire car just to find a bad fuse. (It looked okay but didn't check out with a meter.) That was how he learned this lesson. Me, I just participated in 4 hours of the dumbest argument of my career.

How did the argument end? We never came to an agreement, so I went ahead and fixed boxes with the new design anyway (which they spent several weeks proving were working correctly). I didn't say another word but transferred out of that group as soon as possible. The same design has been in use for over 10 years now, and the documentation notes the need to wire it correctly to avoid inaccurate readings. I didn't write that document, my old boss did. It's kind of funny how we didn't argue about Ohm's Law after that.

The basics are the most important; let me repeat that, the *basics* are *important!* Ohm's Law is the most basic principle you will use as an electrical engineer. It is the foundation on which all other rules are based. The fundamental fact is that resistance impedes current flow. This impedance creates a voltage drop across

the resistor that is proportional to the amount of current flowing through it. If it helps, you can think of a resistor as a current-to-voltage converter.[2]

With that important point made, let's consider two other types of *impedance* that can be found in a circuit. We will get into this in more detail later, but for now consider that inductors and capacitors both can act like resistors, depending on the frequency of the signal. If you take this into account, Ohm's Law still works when applied to these components as well. You could very well rewrite the equation to:

$$V = I * Z \qquad \textbf{Eq. 2.2}$$

Think of the impedance Z as resistance at a given frequency.[3] As we move on to the other basic equations, keep this in mind. Wherever you see resistance in an equation, you can simply replace it with impedance if you consider the frequency of the signal.

One final note: Every wire, trace, component, or material in your circuit has these three components in it: resistance, inductance, and capacitance. Everything has resistance, everything has capacitance, and everything has inductance. The most important question you must ask is, "Is it enough to make a difference?" The fact is, in my own experience, if the shunt resistor had been 100 times larger, that would have made the errors we were seeing 100 times less.[4] They would have been insignificant in comparison to the measurement we were taking. The impedance equations for capacitors and inductors will help you in a similar way. Consider the frequencies you are operating at and ask yourself, "Is this component making a significant impact on what I am looking at?" By reviewing this significance, you will be able to pinpoint the part of the circuit you are looking for.

---

[2]If you don't get the idea of a current-to-voltage converter, think about it a bit harder, put current through a resistor, get a voltage drop across it out… hopefully deep thought on this will lead to one of those "light bulb over the head" moments when it all seems to make sense.

[3]Okay, this is a bit oversimplified; it acts like resistance in one sense, but it does so by causing a delay in the phase of the signal. I have found that in most cases thinking of it like this will give you a decent idea of what is going on. Just remember it isn't exactly like a resistor dependent on frequency; it merely acts like one.

[4]Here's a fun question for you to figure out: If I had used a resistor 100 times larger, what would have been the ramifications of that? What wattage of resistor would I have needed? Would that have affected the operation of the device under measurement? If so, how much, and why? I have found that the brightest engineers will throw a problem like this up on the white board and dig into it, arguing the finer points until their boss comes along and says, "Okay, enough fun, time to get back to work."

The experience I related earlier happened years ago at the beginning of my career, and I said then that I still rediscover Ohm's Law every six months. Time and time again, working through a problem or design, the answer can be found by application of Ohm's Law. So, before you break out all those higher theories trying to solve a problem, first remember: Ohm's Law still works!

The Voltage Divider Rule

Next on our list of basic formulae is the *voltage divider rule*. Here is the equation and Figure 2.5 shows a schematic of the circuit:

$$Vo = Vi \frac{Rg}{Rg + Ri} \qquad \text{Eq. 2.3}$$

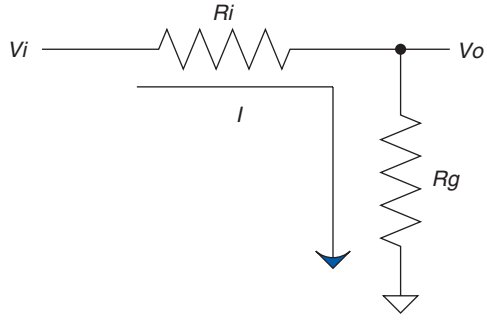


FIGURE 2.5
Input voltage is divided down at the output.

The most common way you will see this is in terms of *R*1 and *R*2. I have changed these to *Rg* (for *R ground*) and *Ri* (for *R input*) to remind myself which one of these goes to ground and which one is in series. If you get them backward, you get the amount of voltage lost across *Ri*, not the amount at the output (which is the voltage across *Rg*). If the gain[5] of this circuit just doesn't seem right, you might have the two values swapped.

You might also notice that the gain of this circuit is never greater than 1. It approaches 1 as *Ri* goes to 0, and it approaches 1 as *Rg* gets very large. (Note that as *Rg* gets larger, the value of *Ri* becomes less significant.) Since this is the case, it is easy to think of the voltage divider as a circuit that passes a percentage of the voltage through to the output. When you look at this circuit, try to

[5]One way I like to think of this is $Vo = Vi * H$, where H is the gain of the circuit, or $H = Rg/(Rg + Ri)$. This is useful when you are breaking a circuit down to components. We will specifically use this when we discuss op-amps later on.

think of it in terms of percentage. For example, if *Rg* = *Ri*, only 50% of the voltage would be present on the output. If you want 10% of the signal, you will need a gain of 1/10. So put 1 K in for *Rg*, and 9 K in for *Ri*, and, voilà, you have a voltage divider that leaves 10% of the signal at the output.

Did you notice that the ratio of the resistors to each other was 1:9 for a gain of 1/10? This is because the denominator is the sum of the two resistor values. I'll also bet you noticed that if you swap the two resistor values you will get a gain of 9/10, or 90%. This should make intuitive sense to you now if you recognize that, for the same amount of current, the voltage drop across a 9 K *Ri* will be nine times larger than the voltage drop across a 1 K *Rg*. In other words, 90% of the voltage is across *Ri*, whereas 10% of the voltage is across *Rg*, where your meter measuring *Vo* is hooked up. The voltage divider is really just an extension of Ohm's Law (go figure), but it is so useful that I've included it as one of the basic equations that you should commit to memory.

Capacitors Impede Changes in Voltage

Let's consider for a moment what might happen to the previous voltage divider circuit if we replace *Rg* with a capacitor. It is still a voltage divider circuit, is it not? But what is the difference? At this point you should say, "Hey, a cap is just a resistor that's value changes depending on the frequency; wouldn't that make this a voltage divider that depends on frequency?" Well, it does, and this is commonly known as an *RC circuit*. Let's draw one now, as shown in Figure 2.6.

Using your intuitive understanding of resistors and capacitors, let's analyze what is going to happen in this circuit. We'll do this by applying a step input. A *step input* is by definition a fast change in voltage. The resistor doesn't care about the change in voltage, but the cap does. This fast change in voltage can be

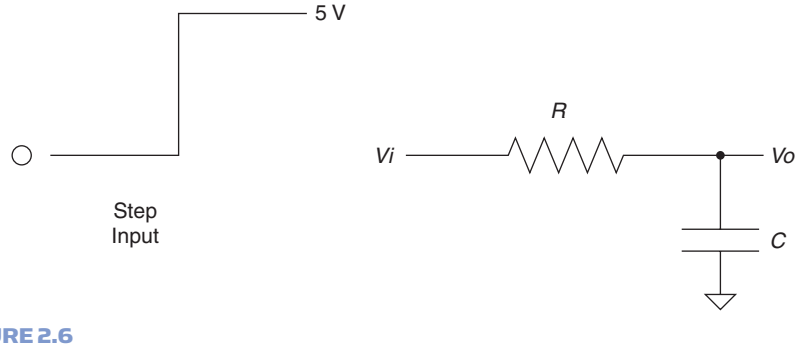


FIGURE 2.6
Step input is applied to a simple RC circuit.

thought of as high frequencies,[6] and how does the cap respond to high frequencies? That's right, it has low impedance. So, now we apply the voltage divider rule. If the impedance of *Rg* is low (as compared to *Ri*), the voltage at *Vo* is low. As frequency drops, the impedance goes up; as the impedance goes up, based on the voltage divider, the output voltage goes up. Where does it all stop?

Think about it a moment. Based on what you know about a cap, it resists a change in voltage. A quick change in voltage is what happened initially. After that our step input remained at 5V, not changing any more. Doesn't it make sense that the cap will eventually charge to 5V and stay there? This phenomenon is known as the *transient response* of an RC circuit. The change in voltage on the output of this circuit has a characteristic curve. It is described by this equation:

$$Vo = Vi\left(1 - e^{-\frac{\tau}{rc}}\right) \qquad \textbf{Eq. 2.4}$$

The graph of this output looks like Figure 2.7. The value of *R* times *C* in this equation is also known as *tau*, or the time constant, often referred to by the Greek letter τ.

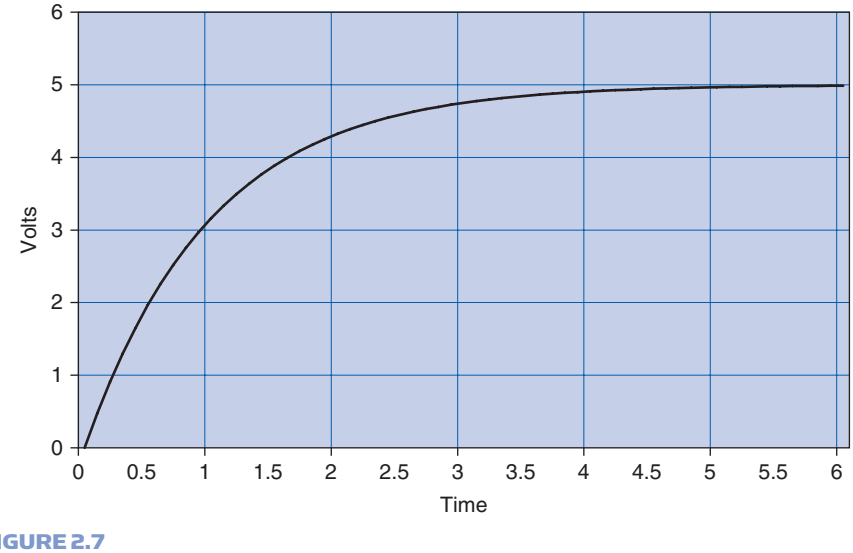


FIGURE 2.7
Voltage change over time.

[6]This is something a man named Fourier thought of long ago. The more harmonic frequencies you sum together, the faster the rise time of said step input.

$$RC = \tau \qquad \text{Eq. 2.5}$$

For a step input, this curve is always the same for an RC circuit. The only thing that changes is the amount of time it takes to get to the final value. The shape of the curve is always the same, but the time it takes to happen depends on the value of the time constant[7] tau. You can normalize this curve in terms of the time constant and the final value of the voltage. Let's redraw the curve with multiples of τ along the time axis, as shown in Figure 2.8.

At 1 τ the voltage reaches 63.2%, at 2 τ it is at 86.6%, 3 τ is 95%, by 4 τ it is at 98%, and when you reach 5 τ you are close enough to 100% to consider it so.

This response curve describes a basic and fundamental principle in electronics. Some years ago I started asking potential job candidates to draw this curve after I gave them the RC circuit shown in Figure 2.6. Over the years I have been dismayed at how many engineers, both fresh out of school and with years of experience, cannot draw this curve. Fewer than 50% of the applicants I have asked can do it. That fact is one of the main reasons I decided to write this book. (The other was that someone was actually willing to pay me to do it! I doubt it would have gotten far otherwise.) So, I implore you to put this to memory once and for

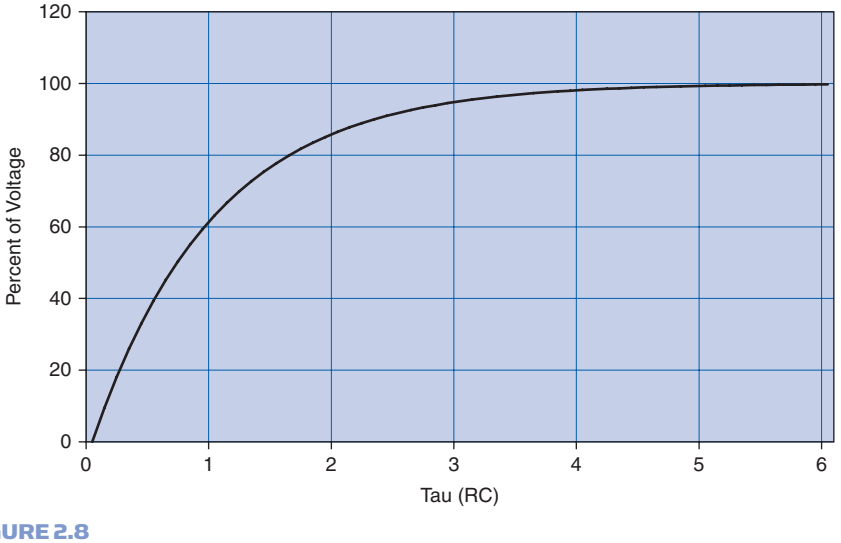


FIGURE 2.8
Voltage change in percent over time in tau.

[7]If you stop to think about it, it just makes sense that this value RC is called a *time constant*, since it affects the timing of the response.

all; by doing so I guarantee you will be a better engineer. Plus, if I ever interview you, you will have a 50% better chance of getting a job! If you understand this concept, you will understand inductors, as you will see in the next section.

Before we move on, I would like you to consider what happens to the current in this circuit. Remember Ohm's Law? Apply it to this example to understand what the current does. We know that:

$$V = I * R \qquad \textbf{Eq. 2.6}$$

A little algebra turns this equation into:

$$I = \frac{V}{R} \qquad \textbf{Eq. 2.7}$$

A little common sense reveals that the voltage across R in this circuit is equal to voltage at the output minus voltage at the input. As an equation, you get:

$$Vr = Vi - Vo \qquad \textbf{Eq. 2.8}$$

We know the voltage at each point in time in terms of tau. At 0 τ, *Vo* is at 0. So the full 5V is across the resistor and the maximum current is flowing. For all intents and purposes, the cap is shorting the output to ground at this point in time. At 1 τ, *Vo* is at 63.2% of *Vi*. That means *Vr* is at 36.8% of *Vi*. Repeat this process, connect the dots, and you get a curve that moves in the opposite direction of the voltage curve, something like what's shown in Figure 2.9.

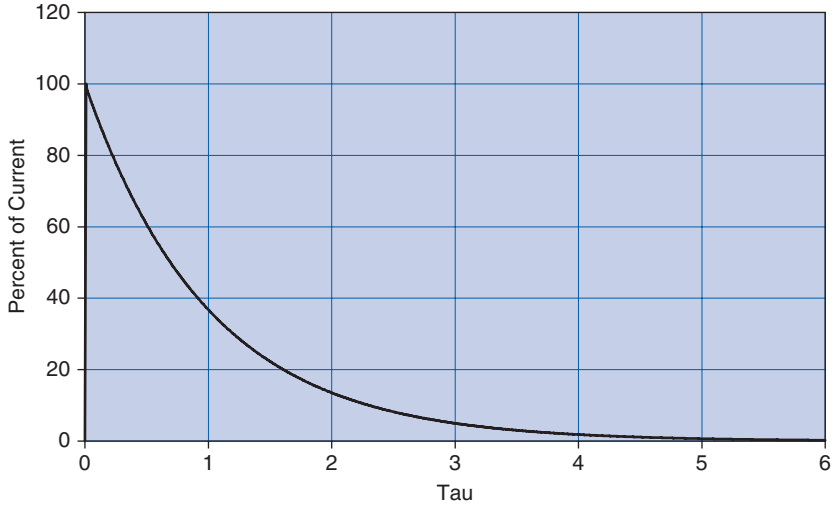


FIGURE 2.9
Current change in percent over time in tau.

Notice how current can change immediately when the step input changes. Also notice how the voltage just doesn't change that fast. Capacitors impede a change in voltage, as the rule goes. What this also means is that changes in current[8] will not be affected at all. Everything has its opposite, and capacitors are no exception, so let's move on to inductors.

Inductors Impede Changes in Current

Now that we have thought through the RC circuit, let's consider the RL circuit shown in Figure 2.10. Remember that the inductor resists a change in current but not in voltage. Initially, with the same step input, the voltage at the output can jump right to 5 V. Current through the inductor is initially at 0, but now there is a voltage drop across it, so current has to start climbing. The current responds in the RL circuit exactly the same way voltage responds in the RC circuit.

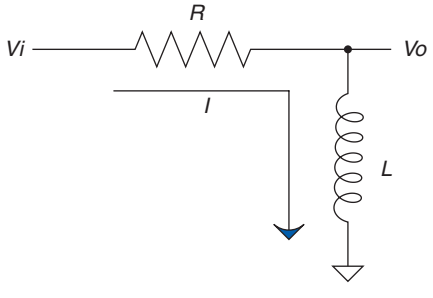


FIGURE 2.10
The basic RL circuit.

Since you committed the RC response to memory, the RL response is easy. It is exactly the same from the viewpoint of current[9]; the current graph looks like Figure 2.11 on the next page.

I hope you are saying to yourself, "What about the voltage response?" At this time, consider Ohm's Law for a moment and try to graph what the voltage will do. What is the current at time 0? How about a little later? Remember Ohm's

[8]Another way to think of it is that rapid changes in current are what capacitors are very good at.

[9]Being able to consider a circuit from either a "voltage" viewpoint or a "current" viewpoint is a valuable skill. Try to formulate an understanding of this concept as you develop your skills in this area.

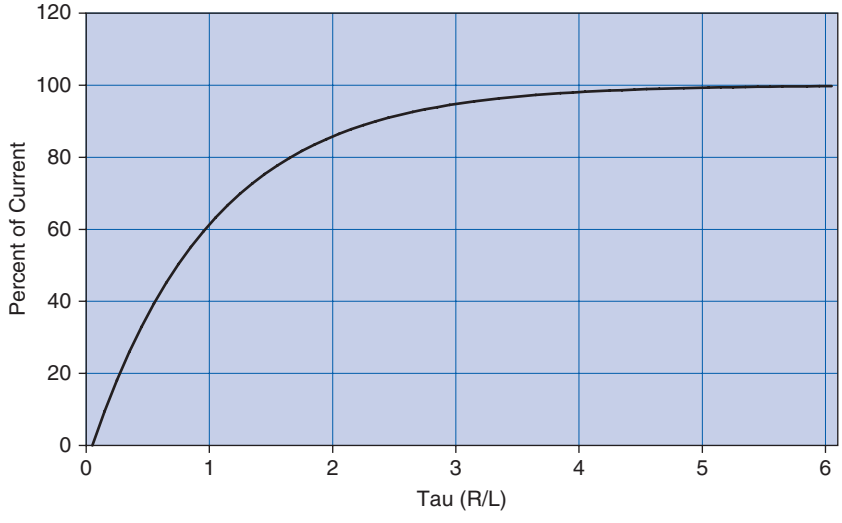


FIGURE 2.11
Current change in percent over time in tau.

Law—for the current to be low, resistance must be high. So initially the inductor acts like an open circuit. Voltage across the inductor will be at the same value as the input. As time goes on, the impedance of the inductor drops off, becoming a short, so voltage drops as well. Figure 2.12 shows the graph.

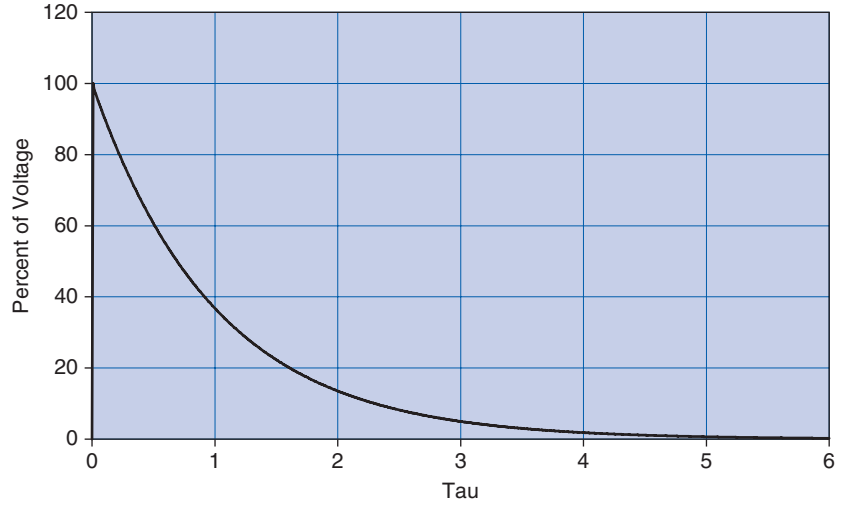


FIGURE 2.12
Voltage change in percent over time in tau.

The inductor is the exact complement of the capacitor. What it does to current, the cap does to voltage, and vice versa.

Series and Parallel Components

There are two ways for components to be configured in a circuit: series and parallel. *Series* components line up one after another; *parallel* components are hooked up next to each other. Let's go over the formulas to simplify these component arrangements.

Series resistors, shown Figure 2.13, are easy; you simply add them up, no multiplication needed!

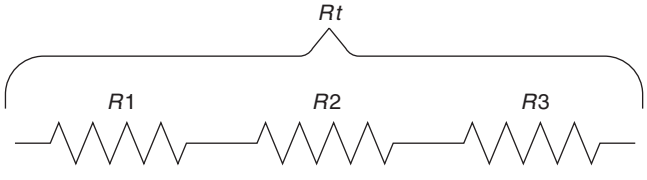


FIGURE 2.13
Series resistors.

$$Rt = R1 + R2 + R3 \qquad \text{Eq. 2.9}$$

The inductors shown in Figure 2.14 are like resistors—you sum series inductors the same way.

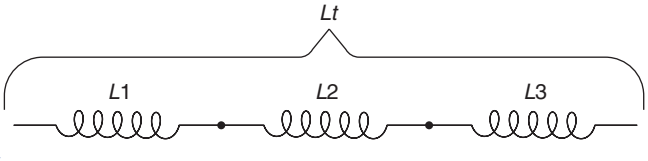


FIGURE 2.14
Series inductors.

$$Lt = L1 + L2 + L3 \qquad \text{Eq. 2.10}$$

Remember that capacitors are the opposite of inductors. For this reason, capacitors must be in parallel to be summed up the way resistors and inductors are in series; see Figure 2.15 on the next page.

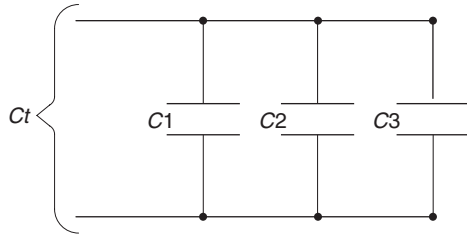


FIGURE 2.15
Parallel capacitors.

$$Ct = C1 + C2 + C3 \qquad \textbf{Eq. 2.11}$$

Remember the equivalences shown in Figure 2.16.

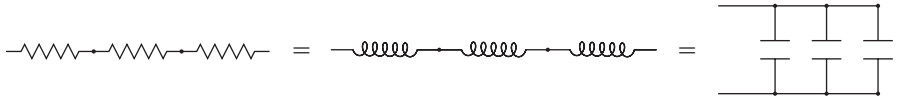


FIGURE 2.16
Component equivalents.

Parallel resistors, shown in Figure 2.17, are a little trickier. The equivalent resistance of any two components is determined by the product of the values divided by the sum of the values.[10]

Keep in mind, however, that this works for any two resistors! In the case of three resistors or more, solve any two and repeat until done. (The // means *in parallel with*.)

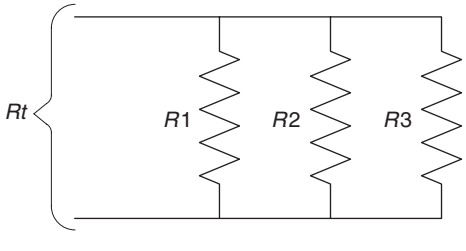


FIGURE 2.17
Parallel resistors.

$$R1 \,//\, R2 = \frac{R1 * R2}{R1 + R2} \qquad Rt = \frac{R1 \,//\, R2 * R3}{R1 \,//\, R2 + R3} \qquad \textbf{Eq. 2.12}$$

$$L1 \,//\, L2 = \frac{L1 * L2}{L1 + L2} \qquad Lt = \frac{L1 \,//\, L2 * L3}{L1 \,//\, L2 + L3} \qquad \textbf{Eq. 2.13}$$

[10]Another way to remember this idea is to sum all the inverses: $Rt = 1/R1 + 1/R2 + 1/R3$, and so on. If this works better for you, that is fine, just commit one or the other to memory.

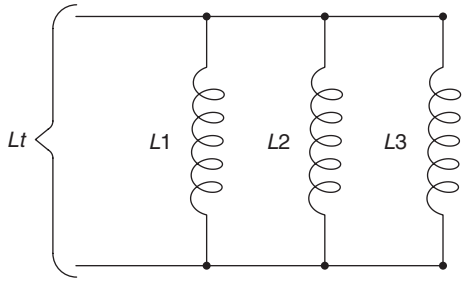


FIGURE 2.18
Parallel inductors.

Parallel inductors are the same as resistors; you can reduce them in the same way—see Figure 2.18.

For capacitors the same equation applies, but only if they are in series, as shown in Figure 2.19. These are the circuits that use the product-over-the-sum, or the sum-of-the-inverses, rule[11]; see Figure 2.20 on the next page.

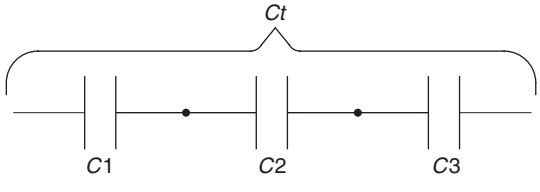


FIGURE 2.19
Series capacitors.

$$C1 // C2 = \frac{C1 * C2}{C1 + C2} \qquad Ct = \frac{C1 // C2 * C3}{C1 // C2 + C3} \qquad \textbf{Eq. 2.14}$$

In dealing with parallel and series circuits, you can see that there are only two types of equations. One is simple addition, and the other is the product over the

[11]You can sum the inverses of the capacitors or inductors in the same way as the resistors. Just put impedance in place of resistance: $Zt = 1/Z1 + 1/Z2 + 1/Z3$, and so on. Truth be told, I committed the product-over-the-sum rule to memory many years ago. That's why I like it. You can be just as effective with the sum of the inverses rule, as so many astute readers have pointed out. I hope you have realized by now that if there are two equivalent routes to get to a destination, I don't particularly care which one you use so long as you get to the right place. I do believe that it is important to find what works for you and focus on that. Don't worry about going a different way unless it gives you new insight or understanding. Wow, this just might be the longest footnote in the whole book! Maybe I should add just a few more words to make sure. If you do read this and make it all the way to the end without nodding off, drop me an email and let me know. Maybe we should have a contest for readers slugging through a footnote. We can put all your names in a hat if you email in, then I will do a drawing and send you an autographed copy of this book! Now there's a reward for reading the tedious parts: more reading!

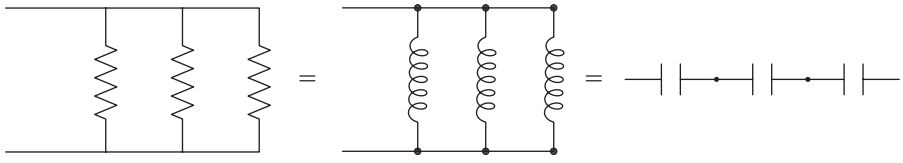


FIGURE 2.20
Component equivalents.

sum (or the sum of inverses). The only trick is to know which to use when. Remember that the resistor and inductor are part of the "in" crowd and the cap is the outcast wallflower who is the opposite of those other guys. I'll bet most engineers can relate to being the "capacitor" at a party, so this shouldn't be too hard to remember!

Thevenin's Theorem

Thevenizing is based on the idea of using *superposition* to analyze a circuit. When you have two different variables affecting an equation, making it difficult to analyze, you can use the technique of superposition to solve the equation, provided that you are dealing with linear equations (by luck all these basic components are linear; even if you might not think it when looking at the curve of an RC time response, it actually is a linear equation).[12]

The idea of superposition is simple: When you have multiple inputs affecting an output, you can analyze the effects of each input independently and add them together when you are all done to see what the output does. One idea that comes from superposition is *Thevenin's Theorem*.

Using Thevenin's Theorem allows you to reduce basically any circuit into a voltage divider. And we know how to solve a voltage divider, don't we! There is a sister theorem called Norton's, which does the same thing but is based on current rather than voltage. Since you can solve any electrical problem with either equation, I suggest you focus on one or the other. Since I like to think in terms of voltage, I prefer Thevenizing a circuit to the Norton equivalent. So to be true to the idea that you should only learn a few fundamentals and learn those well, we will focus on Thevenin equivalents.

[12]When I see the term *linear equation*, I think *line*, so an RC curve seems counterintuitive, but linear equations are a type of formula that allows certain rules such as superposition to be used.

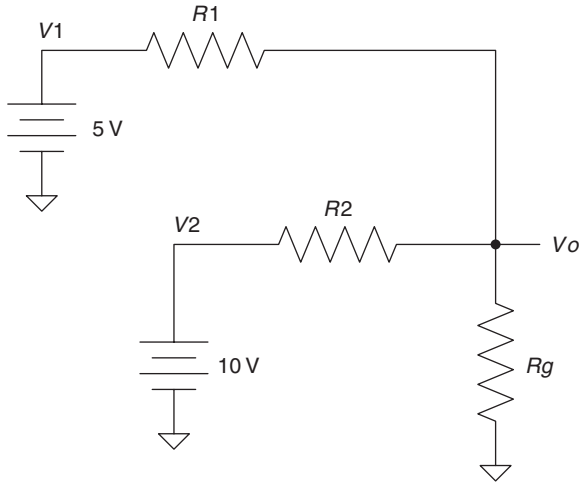


FIGURE 2.21
Circuit with two voltage sources.

The most important rule when Thevenizing is this: *Voltage sources*[13] *are shorted, current sources are opened.* Consider the circuit shown in Figure 2.21.[14]

Once all the voltage sources are shorted and all the current sources are opened, all the components will be in series or parallel. That makes it very convenient for those of us who only want to memorize a few equations! Apply those basic parallel and series rules we just learned and voilà, you have a circuit that is much easier to understand. Once you have reduced the resistors, inductors, and caps to a more controllable number, you replace each source one at a time to see the effects of each source on the component in question.

When you have considered the effects of each source one at a time, you can add them all together to see the overall effect. In this process of Thevenizing the circuit you are superimposing each output on top of the other to get the output of the combined inputs.

---

[13]Note the use of the word *source*; a voltage source is a device that keeps the voltage constant as a load varies. A current source keeps the current constant in the face of a changing load.

[14]This circuit sparked a competition in the last edition of this book (since I didn't bother to include a solution). Rather than try to show you how smart I am by including a solution in this edition, I am going to change a few values (or maybe not) and invite you as readers to solve this circuit. Drop me a note with your solution; I promise you will win an all-expense-paid email back from me congratulating you on your engineering prowess!

I find it helps when Thevenizing a circuit to try to imagine that you are looking back into the circuit from the output. This means that you imagine what the circuit looks like in terms of the output. We often think in terms of stuff that goes in the input. Something goes in, something happens, and then it comes out the output. Try flipping that notion on its head. Think, "Here is the output, what exactly is it hooked up to? What are the impedances that the cap in this case 'sees' connected to it?" Once you are able to adjust your point of view, Thevenizing will become an even more powerful tool. Consider the circuit shown in Figure 2.22.

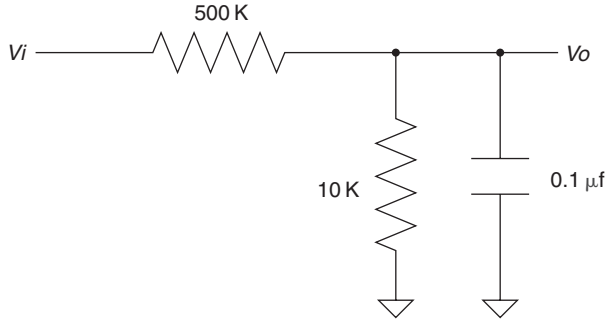


FIGURE 2.22
Real, live secret circuit.

This circuit comes from a real live application. I would tell you what, but it is secret.[15] So we won't say this is anything more than a voltage divider with a capacitive filter on it. (Values might have been changed to protect the innocent.)

This circuit's job is to lower a voltage at the input terminals varying from 0 to 100 V to something with a range of 0 to 5 V. The input voltage also has an AC component that is filtered out by the capacitor. The question is, what is the time constant of the RC filter in this circuit?

Is it 500 K * 0.1 µf? That's what I would have thought before I understood Thevenin's Theorem. The output in this case is the voltage across the cap, so let's look back into the circuit to figure out what is hooked up to this cap. Now remember, I said there was a voltage source on the input of this circuit. Let's

[15] If you haven't already, you will soon find out that every corporation wants you to sign away every idea you have or ever had as their intellectual property. Some day, those individuals who have all the good ideas must rise up and say, "Enough is enough!" After which we will all likely end up being consultants.

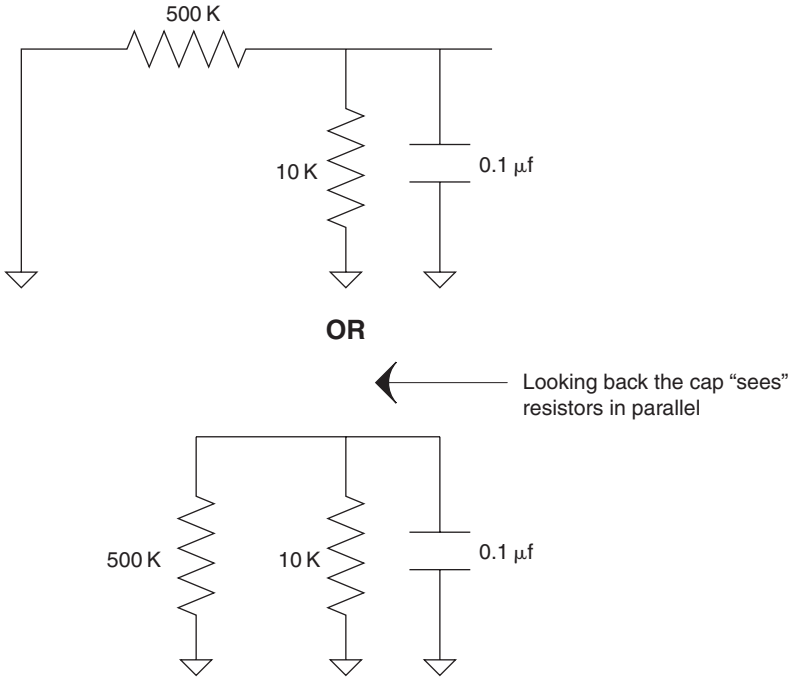


FIGURE 2.23
Thevenized real live secret circuit.

short that on our drawing and Thevenize it. Take a look at the Thevenized circuit shown in Figure 2.23.

Hopefully at this point something really jumped out at you. The 10 K and the 500 K resistors are in parallel as far as the cap is concerned. Applying the rule of parallel resistors we find that the resistance hooked up to this cap is 9.8 K. Wow, that is a lot less than 500 K, isn't it! Thevenizing showed us that our first assumption was incorrect. In fact, the time constant[16] of this circuit is much, much lower than it would be without the 10 K resistor.

There are other ways this theorem can be useful. Here is a case in point. You might have a circuit like the one shown in Figure 2.24 on the next page.

You need to switch AC power through this inductor (which was actually one winding of an AC motor in this case). Trouble is, when you let off the switch, a whole

[16]If you want a more in-depth lesson on time constants, you will need to jump ahead a few chapters. For now, though, it is sufficient to understand the basic idea behind good ol' Thevenin's proposition.

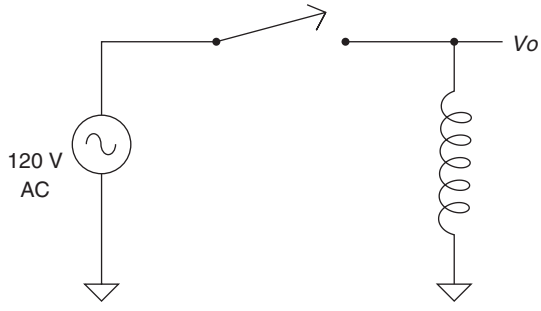


FIGURE 2.24
AC switched power to an inductor.

bunch of electrical noise is generated when this switch is opened. (We will discuss why when we cover magnetic fields later in the book.) A standard way to deal with this situation is with an RC circuit commonly known as a *snubber*. The point of a snubber is to snub this voltage spike and dissipate it as heat on the resistor. This makes the most sense if it is across the inductor, as shown in Figure 2.25.

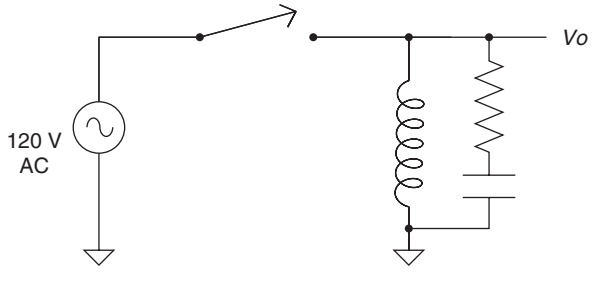


FIGURE 2.25
AC switched power to an inductor with snubber.

Now let's apply Thevenin's Theorem to take a different look at this circuit, as shown in Figure 2.26. By shorting the AC voltage source, we quickly see that hooking the snubber up to the other side of the switch, to the AC hot line, would have exactly the same effect as hooking across the inductor. This fact once saved a company I worked for tens of thousands of dollars[17] using the alternate location of the snubber circuit. I would say that makes Thevenizing a pretty powerful tool, wouldn't you?

[17]No, I didn't get any bonus for my discovery and work in this case. Alas, that too is a sad fact of the corporate world we live and work in these days. On the flip side, the corporate world has made cartoons like *Dilbert* quite successful. Someday I predict corporations will come to realize that the best way to keep people happy while working hard is simply slipping them a few extra $$.

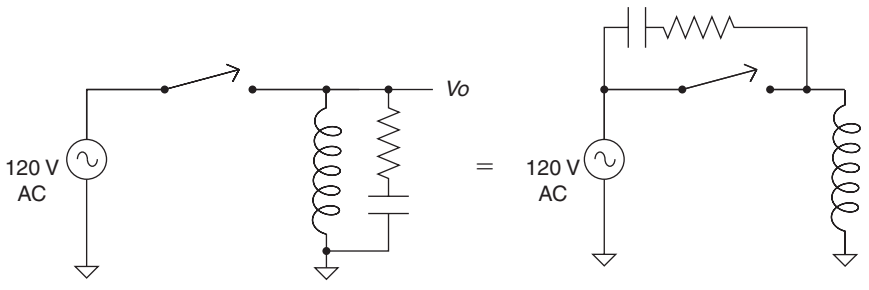


FIGURE 2.26
These are equivalent circuits when Thevenized.

Thumb Rules

- The basics are the most important!
- For a basic understanding, think of impedance as similar to resistance at a given frequency.
- $V = I * Z$.
- Voltage divider rule, $Vo = Vi(Rg/(Rg + Rs))$.
- A capacitor resists a change in voltage, but current can change immediately (the inverse of the inductor).
- An inductor resists a change in current, but voltage can change immediately (the inverse of the capacitor).
- A capacitor is to voltage as an inductor is to current.
- Series resistors, series inductors, and parallel caps add up.
- Parallel resistors, parallel inductors, and series caps use the product-over-the-sums, or the sum-of-the-inverse, rule.
- When Thevenizing: short voltage sources, open current sources.
- Consider the circuit from the output point of view.
- Insight can be gained by Thevenizing a circuit.

IT'S ABOUT TIME

AC/DC and a Dirty Little Secret

AC/DC—it isn't a rock band; it's one of those lovable engineering acronyms. It means *alternating current* and *direct current*. These terms came into being to describe a couple of different modes of electricity. A firm understanding of these two modes will help you in all aspects of engineering. Before we move

on to these two modes, we need to establish an understanding of something called *conventional flow* and *electron flow*.

Way back before we even knew that electrons existed, electricity was thought of as a flow of something. Benjamin Franklin picked a direction for that flow, labeling one side positive and the other negative. (The reason is a whole other story involving wax, wool, and a lot of rubbing.[18]) The presumptions made sense, but it turned out later, as we came to understand what electrons are, that electricity wasn't really a flow and that the electrons actually move in the other direction.

The truth is that the little electrons that produce what we call electricity aren't really a continuous flow of juice; they sort of bump around in these little packets (we learned to call them charges in Chapter 0). From an aggregate level, though, these packets of quanta[19] act like a flow. It also turned out that these charges moved in the opposite direction of what was previously assumed.

By the time all this was figured out, the conventional flow positive-to-negative nomenclature was pretty well established. Since all the basic equations work either way, no one has bothered to change this idea. Instead another term, *electron flow*, is used to describe the way electrons actually move in a circuit.

This seemed like a dirty little secret to me when I first found it out, and I've often wondered whether we haven't missed an important discovery along the way due to thinking of electricity in the manner that we do. Considering the world to be flat doesn't cause significant errors with geometry till you are trying to fly a plane to China and discover that what you thought was a straight line is really a curve. So, as long as we keep things in perspective, the accepted jargon will do fine.[20] For our discussion, we will use conventional flow terminology. We will also consider the effects from an aggregate level, preserving the idea of flow.

Now that you know the dirty little secret of electron flow, let's talk about current and voltage and where they come from.

---

[18]When you have an evening with nothing better to do, I suggest you Google this topic; it is very interesting.

[19]Ahh, quantum mechanics, an interesting and entirely other topic that we will have to reserve for another book at another time.

[20]Don't let that stop you from wondering, though; maybe you will discover something new!

Constant Voltage Sources vs. Constant Current Sources

Devices that cause electrons to move are called *sources* since they are the source of electron or charge flow. The two typical types of source are voltage and current sources. Remember two important things when dealing with them.

IMPORTANT THING 1

When you're dealing with a voltage source, the output will try to maintain the voltage across the load. That is, the voltage at the source will be constant. This means in terms of Ohm's Law that V remains the same at the source. I and R can change, but in the end it must always equal V in terms of Ohm's Law. $V = IR$.

IMPORTANT THING 2

When you're dealing with a current source, the output will try to maintain the current through the load. That is, the current from the source will be constant. These are less common but they do exist and can be used in many situations. Current from the source will remain constant, allowing V to change as R varies, still following Ohm's Law as obediently as any other circuit. $I = V/R$.

The world of electronics is very voltage-centric, so you will see voltage sources much more often than current sources. This being the case, I will concentrate more on these types of sources.

Sources can come in two different types: AC or DC. Let's take a closer look.

Direct Current

The term *direct current* is used to describe current that flows in only one direction. I think this makes direct current the simplest to understand, so we should start there.

Direct current moves only one way, from positive to negative.[21] A battery is a common direct-current device. Hooked up to a load such as a resistor, the current will go something like what's shown in Figure 2.27 on the next page.

A battery[22] is also a constant voltage device, so it will apply whatever current is needed to maintain its output voltage. So, we have 12 V hooked up to 1 Ω of resistance—hey, we just learned how to figure out current on a circuit like that! (More scribbling on a napkin…) That would be 12 amps of current.

[21] Conventional flow considered here.

[22] You can think of a battery as a chemically powered version of the "electron pump" discussed in Chapter 0.

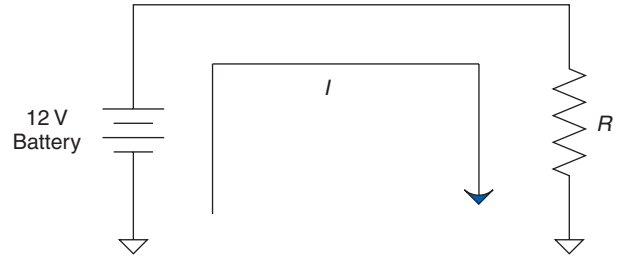


FIGURE 2.27
DC current and voltage from a battery.

A DC source will always try to move current in the same direction. One thing to note is that the current coming out of the source always needs to get back to the source somehow. The ground connection on the schematic should be thought of as a label that connects the signal back to the source. If the signal does not get back to the source, then there is no current flow.[23]

Alternating Current

AC or alternating current came about as the interaction of magnets and electricity were discovered. In an AC circuit, the current repetitively changes direction every so often. That means current increases in flow to a peak, then decreases to zero current flow, then increases in flow in the opposite direction to a peak, then back to zero, and the whole process repeats. The current alternates the direction of flow in a sinusoidal fashion, so of course it is called *alternating current*. This type of current most commonly comes from big AC generators at your local hydroelectric dam.

AC power came into being due to this ease of generation—see Figure 2.28. When you move a coil of wire past a magnet, the current first climbs as the

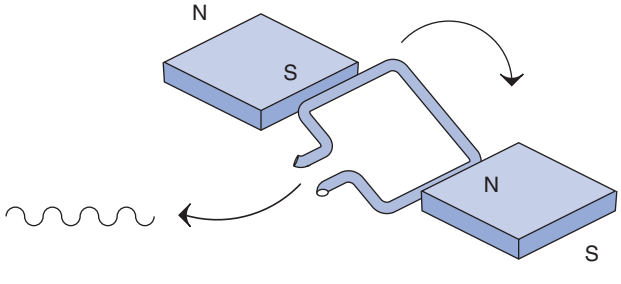


FIGURE 2.28
Simple AC generator.

[23]There are those that would argue this point. If you want to know more, do a search for *free energy* on the Internet, but beware—much of it is complete bunk. That doesn't make it a bad read, though; it can be humorous and quite thought provoking.

strength of the field increases, then as the field decreases and switches polarity, the current also decreases and switches polarity. The voltage and current change in a sinusoidal fashion naturally as the coil passes by the magnets.

As long as you keep moving the coil, AC power will continue to be generated. You will see an AC source on a schematic represented by a sine wave squiggle like the one shown in Figure 2.29.

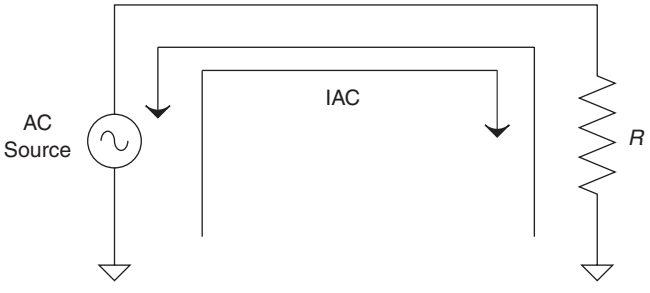


FIGURE 2.29
AC voltage and current source.

An interesting side note is that there was some argument when plans were being drawn up to distribute electricity across the United States. Edison (yeah, the famous light bulb guy) wanted to put small DC generators in everybody's home. Another lesser-known genius by the name of Tesla was pitching for AC distribution by wires from a central location. AC made some sense because the voltage could be easily transformed (yeah, you guessed it, with a transformer) from one level to another. That made it possible to jack up the voltage so high that the resistance of the distribution wires had little loss over long distances. There was much debate over the best setup.

One thing that tipped the scales in the direction we have today was the invention of the AC motor by Tesla. Till then only DC motors had been developed, and since this was before the diode, it wasn't so easy to make AC into DC. So being able to run a motor was a big deal. Although not as famous as Edison, Tesla[24] left a huge legacy in terms of AC power distributions and AC motors. Just look around your house and count up the AC motors in use. (Of course, there are a few light bulbs around, too.)

[24]Nikola Tesla (1856–1943) actually signed over patents for AC power distribution that years later were worth trillions of dollars. If he hadn't done so, our power systems might have been very different than they are today. To learn more about the somewhat sad story of this genius, I suggest reading *Tesla: Man Out of Time*, by Margaret Cheney.

Back to Capacitors and Inductors Again

What was the rule of thumb for a capacitor? Capacitors impede a change in voltage. Do you remember the rule for inductors? Inductors impede a change in current. The flip side of these rules is that the cap will let current change all it wants and the inductor will let voltage change all it wants. One fact you can't ignore when it comes to AC sources is that current and voltage are always changing. How fast they change is a function of a term known as *frequency*. Frequency is the number of cycles of change per second; it has a unit called *hertz*. The higher the frequency, the faster the change in voltage and current. Now extrapolate for a moment what might happen to a cap in an AC circuit. It follows that a cap will block currents that have zero frequency (like a DC battery) and pass currents that change. The opposite applies to an inductor.

I like to think of it this way: The cap is an infinite resistor at DC or zero frequency. As frequency increases, the "resistance" (technically known as *reactance*) of the cap gets lower and lower, approaching zero. This capacitive reactance is known as X_C and is described by this equation. The unit is ohms, just like a resistor.

$$X_c = \frac{1}{2\pi * f * C} \qquad \textbf{Eq. 2.15}$$

The inductor is just the opposite. It starts with $0\,\Omega$ of resistance at a zero frequency and then increases to infinity along with the frequency. Inductive reactance is called X_L and follows this equation:

$$X_L = 2\pi * f * L \qquad \textbf{Eq. 2.16}$$

Let's hook them up to an AC source and vary the frequency to see how current flow is affected. This is easy enough to do in a spreadsheet. Simply plug the reactance equation into Ohm's Law. The schematic is shown in Figure 2.30.

Figure 2.31 shows what happens to current as frequency varies from zero (DC) to really fast (AC) when hooked up to a voltage source.

So to repeat, the higher the frequency, the easier the current wisll pass through a capacitor and the harder it becomes for it to get through an inductor.

You might be thinking, "What about that step input we put into the RC circuit before? How is it AC?" Actually, as weird as it might sound, it is AC. A really smart guy by the name of Fourier figured out some time ago that hidden in fast-changing signals are all sorts of high frequencies. He proved that the more abrupt the change in the signal, the more high frequencies that are present. An

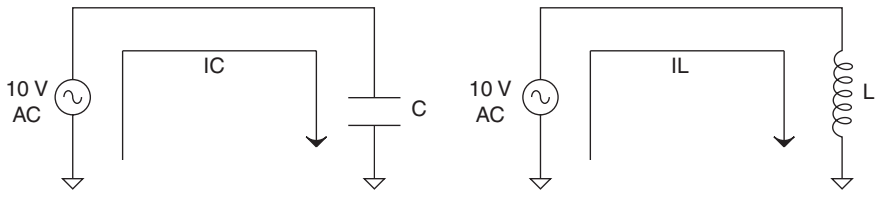


FIGURE 2.30
AC source hooked up to cap and inductor.

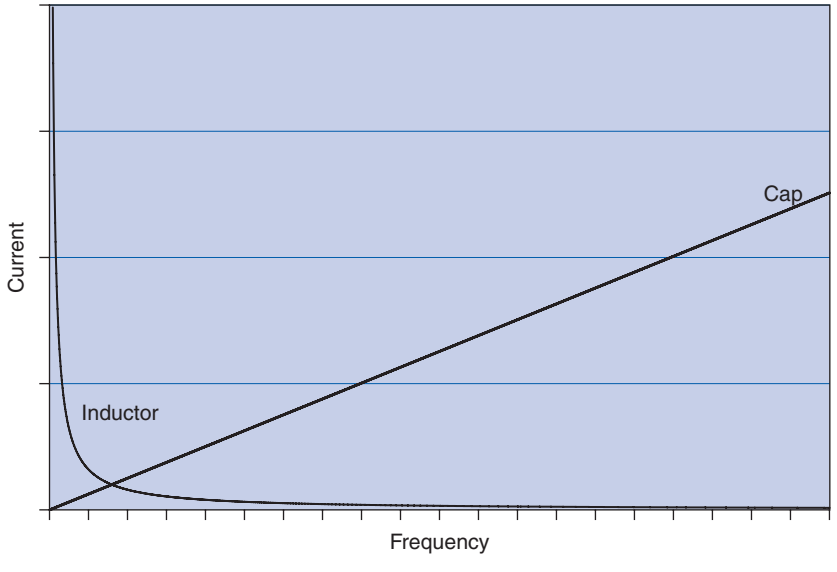


FIGURE 2.31
Graph of current over frequency for a cap and an inductor.

in-depth study of this topic is somewhat beyond the scope of this book, so let it suffice to say that the step input in the previous discussion has a sharp square corner, hidden in which are a whole bunch of high frequencies. These can't get through the cap, so the corner is knocked off, so to speak, leaving the characteristic curve that we saw as the transient response of the RC circuit.

Before we move on, we should touch on the topic of *phase shifts*. When both voltage and current are in sync, they are *in phase*. As we have discussed numerous times, inductors impede a change in current, but voltage is not affected, so if you graph the relationship between voltage and current, you will see that the change in current is a little out of sync with the change in voltage. It is said to be lagging behind. The capacitor has the opposite effect (as always),

so the voltage is delayed relative to the current. In this case the change in current leads[25] the change in voltage. The current isn't magically jumping ahead of the voltage; the voltage is getting behind, but from the voltage point of view it looks like the current is changing first.

Capacitors and inductors are components that impede[26] a signal, the amount of which depends on the frequency of the signal. The cap delays voltage changes, and the inductor delays current changes. They are opposite in the way they react to the frequency of a signal. Caps block lower frequencies while letting higher ones through, whereas inductors pass lower frequencies while blocking higher ones. Let's see what happens when we hook them up to a resistor.

Low-Pass Filters

Consider the circuit shown in Figure 2.32. Note similarities to the RC circuit that we used to first understand the effects of a capacitor. The difference is that now we are going to apply an AC signal to the input rather than the step input we applied before.

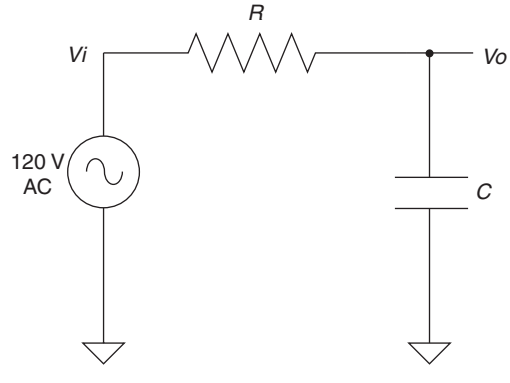


FIGURE 2.32
Cap-based low-pass filter.

[25] I may just be a slow learner, but it took me a while to understand this leading vs. lagging terminology and what it really means. There isn't a magical time machine in a cap that makes the current change before the voltage; it is merely delaying the current change. Of course, if there were a time machine in there, we'd have to call it a "flux" capacitor!

[26] Like resistance, but not exactly; keep in mind that it is an analogy, a very close one, but an analogy none the less. This behavior is the result of the phase delay we discussed earlier. Think way back to Chapter 1: The mass doesn't have friction, but it feels the same as friction when you start to move it. If you were to move it back and forth at the right rate, you could get it to feel exactly like friction. The shifting phases in voltage and current create the same resistance effect when dealing with caps and inductors.

This circuit is known as a *low-pass filter*, and all you really need to know to understand it is the voltage-divider rule and how a capacitor reacts to frequency. If this were a simple voltage divider, you could figure out, based on the ratio of the resistors, how much voltage would appear at the output. Remember that the cap is like a resistor that depends on frequency and try to extrapolate what will happen as frequency sweeps from zero to infinity.

At low frequencies the cap doesn't pass much current, so the signal isn't affected much. As frequency increases, the cap will pass more and more current, shorting the output of the resistor to ground and dividing the output voltage to smaller and smaller levels. There is a magic point at which the output is half the input.[27] It is when the frequency equals $1/RC$. You might have noticed that this is the inverse of the time constant that we used earlier when we first looked at caps. Kinda cool when it all comes together, isn't it?

This is known as a *low-pass filter* because it passes low frequencies while reducing or attenuating high frequencies. You can make a low-pass filter with an inductor and resistor, too. Given that the inductor behaves in a way that is opposite of a capacitor, can you imagine what that might look like? Have a look at Figure 2.33.

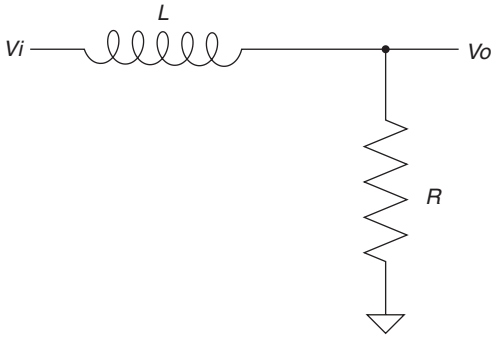


FIGURE 2.33
Inductor-based low-pass filter.

That's right; you swap the position of the components. That's because the inductor (being the opposite of a cap) passes the lower frequencies and blocks the higher frequencies. It performs the same function as the low-pass RC circuit

[27]This is also known as the *–3 db down point*. I am avoiding decibels for now to limit the amount of knowledge that you need to assimilate.

but in a slightly different manner. You still have a voltage-divider circuit, but instead of the resistor-to-ground changing, the input resistor is changing. At low frequencies the inductor is a short, making the ground resistor of little effect. As frequencies increase, the inductor chokes[28] off the current, reacting in a way that makes the input element of the voltage divider seem like an increasingly large resistance. This in turn makes the resistor to ground have a much bigger say in the ratio of the voltage-divider circuit.

To summarize, in the low-pass filter circuits, as the frequencies sweep from low to high, the cap starts out as an open and moves to a short while the inductor starts out as a short and becomes an open. By positioning these components in opposite locations in the voltage-divider circuit, you create the same filtering effect. The ratio of the voltage divider in both types of filters decreases the output voltage as frequencies increase. All this lets the low frequencies pass and blocks the high frequencies. Now, what do you suspect might happen if we swap the position of the components in these circuits?

High-Pass Filters

Swapping the cap and the resistor in the low-pass circuit creates another type of circuit called a *high-pass filter*. Using your now supreme powers of deduction and intuition, you are thinking to yourself, "I'll bet that means the circuit passes high frequencies while blocking low ones." You are correct, and the circuit looks like the one in Figure 2.34.

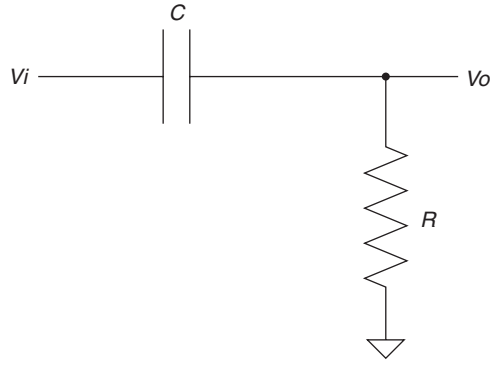


FIGURE 2.34
Cap-based high-pass filter.

[28]Without any proof whatsoever, I assert that inductors are sometimes called *chokes* for the reason that they choke off high frequencies.

Hopefully, after our discussion on the low-pass circuit, the operation of this one is clear. The cap acts like a larger resistor at low frequencies, making the voltage divider knock down the output. At higher frequencies the cap passes more current as it becomes a short, causing a higher voltage at the output. The inductor version of this circuit looks like Figure 2.35.

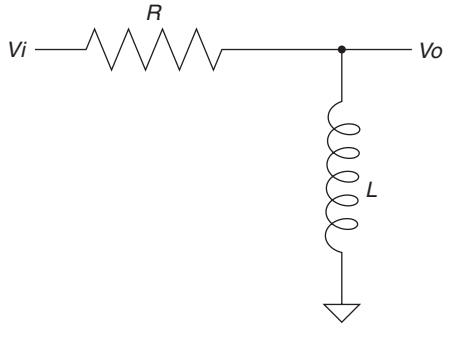


FIGURE 2.35
Inductor-based high-pass filter.

As you might have suspected, this filter is the inverse, circuit-wise, of the RC high-pass filter. Another little bit of serendipity is the fact that the half-voltage output point[29] is also at 1/tau (*tau* means time constant generically, whether referring to an RC or an RL circuit), just like the low-pass filters.

To sum up, the high-pass and low-pass filters take advantage of the frequency response of either a capacitor or an inductor. This is done by combining them with a resistor to create a voltage divider that attenuates the unwanted frequencies while allowing the desired ones to pass. Some cool things happen when we put the two reactive elements together. You can create notch and band-pass filters where a specific band of frequencies is knocked out, or a specific band is passed while all others are blocked. The phenomenon of resonance also occurs in what is called a *tank circuit*, where you have a capacitor combined with an inductor. Just like the spring-mass example in Chapter 1 the tank circuit will oscillate current back and forth from one component to the other.

Active Filters

So far we have been studying passive filters. A passive component is one that is not powered externally. Being passive, these components are subject to an effect known as loading. This means that anything you hook up to the output

[29]This point is also known as the *rolloff* or *3 db down point*.

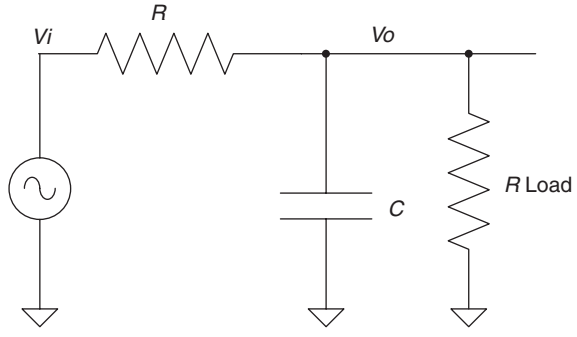


FIGURE 2.36
Filter with load.

can affect the performance of the filter. Take a low-pass RC filter, for example, and hook a resistor up to it, as shown in Figure 2.36.

This resistor on the output is a load. It could be another part of the circuit or any number of things, but the point is that it acts like a resistor to ground. How does this affect the RC filter performance? To understand, let's Thevenize it to "see" how the load affects the output. We start by shorting the voltage source to ground. This is done with AC sources the same as DC, so the circuit would look like Figure 2.37.

Because I like my examples to use real numbers,[30] let's make up some values. Let $R = 10\,\text{K}$ and let $Rload = 10\,\text{K}$ and $C = 0.1\,\mu\text{f}$.

When you Thevenize a circuit, you reduce all the parts into one, where possible. In this case the resistors are in parallel, so apply the parallel rule to the resistors and

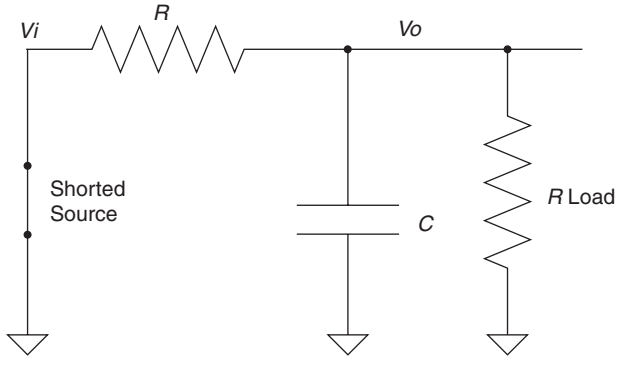


FIGURE 2.37
Thevenized circuit shows effect of load.

[30]To understand how something works, I suggest plugging in real numbers. Even when your final goal is an algebraic equation, running some real numbers through will help you get a genuine understanding of how the thing is really working.

you get a value of 5 KΩ. Did you notice that the *R* value has changed considerably due to the load on the circuit? What might seem counterintuitive at first is the fact that the time constant of this circuit is a function of the Thevenized version that we just derived. So, without the load, tau would have been 10 K * 0.1 μs, or 1 ms.

With the load, it is 0.5 ms, half what it was before! Since the output of this filter depends on τ, we can see that the load has affected it significantly. A way to avoid this problem is to add an active component to the design, making it into an "active" filter. In adding such a component, the basic idea is to minimize this loading effect to a point that you get a nice predictable response. The output of the active filter is such that no matter what load you put on it, it does not affect the response of the filter—see Figure 2.38.

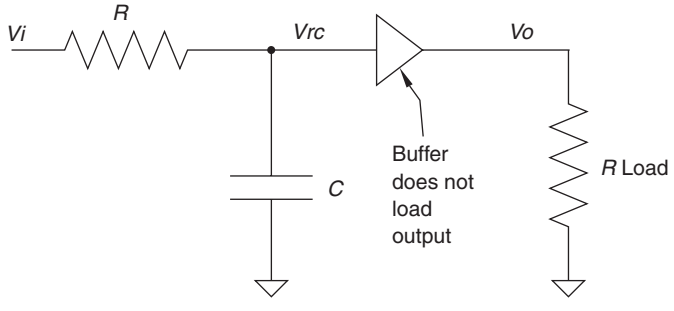


FIGURE 2.38
Active buffer eliminates the effect of the load.

The input of this active device (known as an *op-amp*) has a very high impedance. In this case it is comparable to a 10-meg resistor. Hooking that up to the RC filter will have little effect on the time constant of this circuit as long is it is significantly larger[31] than the *R* value in the circuit. The buffer in this circuit will output a voltage that matches the voltage on the input. It will *buffer* the signal; no matter what you hook up to the output, the filter will not be affected. This is one of the simplest active filters, but the principle with all of them is the same—include an active element to preserve or enhance the integrity of the filter.

[31]Here is a good place to use those estimation skills that we discussed back in Chapter 1. Particularly ratios, if the load on the circuit is 100 times larger (or a 100:1 ratio) than the resistance used in the filter, you can see that the effect of this device won't do much to the circuit it is hooked up to.

Thumb Rules

- Electrons move from negative to positive.
- Direct current flows in one direction; it has a zero frequency component.
- Alternating current changes direction of flow repetitively.
- Direct current has a frequency component of zero.
- Inductors and capacitors can make both low-pass and high-pass filters when combined with a resistor.
- Inductors and caps in the same circuit will oscillate.
- Active filters add components to preserve or enhance the integrity of the filter.

BEAM ME UP?

Electrons are everywhere, or maybe it is better said that the *effects* of electrons are everywhere. There are invisible fields of force all around us that are caused by those pesky little devils. These fields warrant discussion because they are a factor in the way we deal with electricity. They can store energy and affect the world around them in various ways, so it is good to build an intimate knowledge of these fields and how they interact. These are the same fields that we talked about in the newly added Chapter 0 of this edition. If you elected to skip it when you started reading and you feel a little lost here, I suggest going back to read that chapter.

Now, we can't beam people[32] on and off the planet like they do on *Star Trek*, but there are couple of invisible fields that are what actually create all the effects that we see in electric circuits. It is these fields that make the water, ping-pong ball, and many other analogies not quite match what is really going on in a circuit. So understanding these fields will definitely help develop the intuitive skills we are working on.

The Magnetic Field

This is the most well-known of the two fields that we are going to discuss. Who hasn't experienced the force of a magnet sticking a note to the fridge or felt

[32]Note I said *people*; physicists have "beamed" a quantum bit from one point to another. Some say teleporting an atom could happen in just a few years.

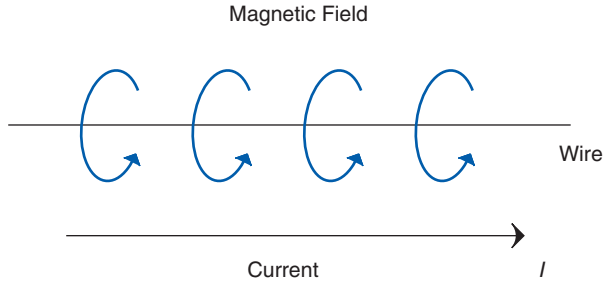


FIGURE 2.39
Magnetic field caused by current in a wire.

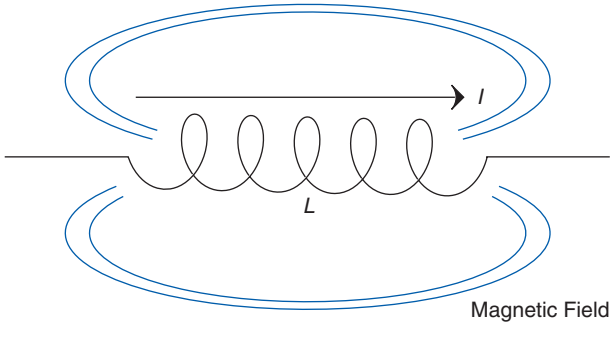


FIGURE 2.40
Coils change direction of field and reinforce.

the power of two repelling magnets? Back in the 1820s, a man by the name of Hans Oersted noticed his compass read strangely every time he switched on a current in a wire. Eventually it was figured out that a moving electron (such as the current in a wire) creates a magnetic field perpendicular to the direction of electron movement, as shown in Figure 2.39.

This field is identical to the field surrounding a permanent magnet. In fact, if you coil the wire like what's shown in Figure 2.40, the magnetic field lines align and reinforce each other, making it even more like a permanent magnet.[33]

Electromagnets, as they are called, are pretty cool since they can be switched on and off, unlike permanent magnets. Another important fact is that not only

[33]As mentioned earlier in Chapter 0. In a permanent magnet you have a whole bunch of electrons spinning around in the same direction. This is kinda like the way the flow in the wires reinforces the field using a coil of wire.

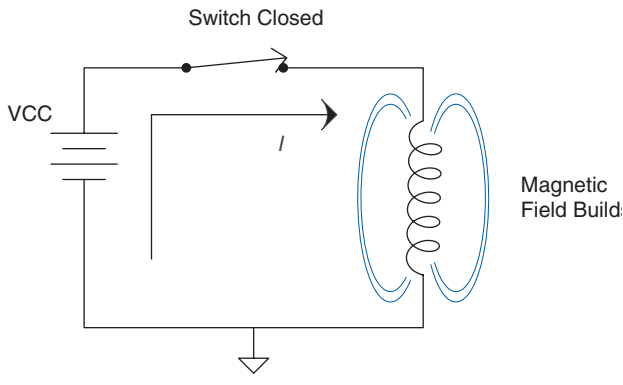


FIGURE 2.41
Building a magnetic field resists current change.

does a current moving through a wire create a magnetic field, but the opposite is also true. A changing magnetic field can create or induce a current in a wire. A coil of wire is known as an *inductor* for this reason. Energy is stored in an inductor as a magnetic field. It is like a rubber band that is stretched as you apply current. When the current is shut off it snaps back, and energy is given up as the magnetic field collapses (it is changing as it goes away). This collapse induces a current in the wire. Consider the circuit shown in Figure 2.41.

When the switch is closed, current flows and a magnetic field is created. It is the creation of the magnetic field (stretching the rubber band, so to speak) that causes the inductor to "resist"[34] the change in current, as we learned it does earlier. The flip side of that also happens. If we open the switch, the change in the field as it collapses would like to keep the current flowing in the inductor—see Figure 2.42.

If there is no place for this current to go, the voltage across the inductor will increase instantaneously and then dissipate as the induced current drops off with the drop of the magnetic field. Take a look at the graph shown in Figure 2.43 of the current and voltage changing in this inductor circuit as the switch opens and closes.

[34]Here is where the reactive components (caps and inductors) aren't exactly like a frequency-dependent resistor. The creation of these fields takes energy, and the dissipation of these fields releases this stored energy. In a resistor this energy is converted to heat; not so in a cap or inductor (albeit a perfect one). In these components the storing and releasing of energy is what causes them to seem like a resistor as they "impede" current or voltage changes in a circuit.

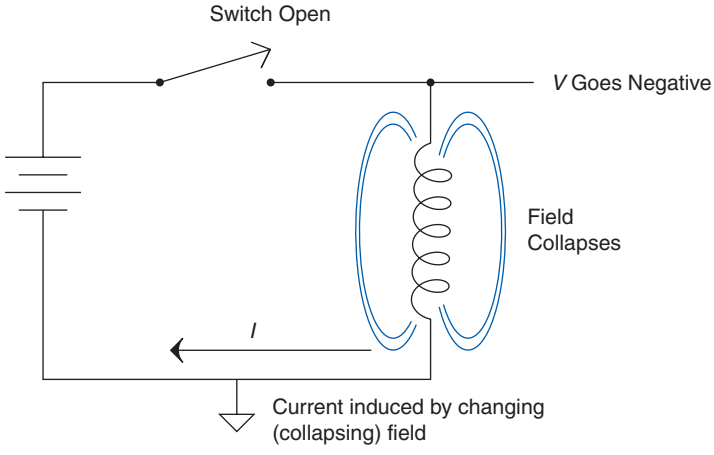


FIGURE 2.42
Collapsing magnetic field generates a current.

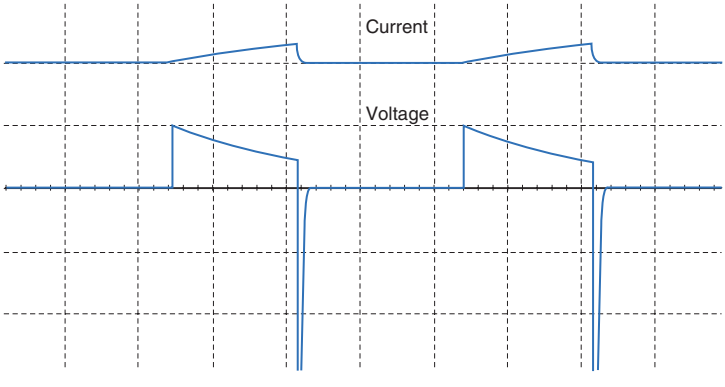


FIGURE 2.43
Voltage and current changes as an inductor is switched in and out of circuit.

Induction is also the fundamental principle that a transformer uses. The magnetic field—as it is created on one side of the transformer as is shown in Figure 2.44—induces a current on the other side of the transformer. When the field reduces, and it switches direction, a corresponding current is induced at the output.

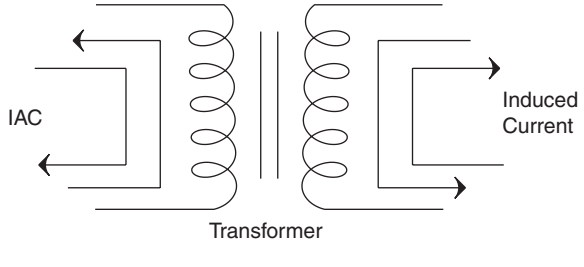


FIGURE 2.44
A transformer uses changing current on the input to induce current on the output.

The ratio of turns on each side of the transformer controls the ratio of voltage from input to output. A 10:1 ratio will take 120 V on one side and create 12 V on the other. Note also that though voltage goes down, current goes up, making a transformer kind of like a gear train or lever in the mechanical world. Power into it is the same as power out of it (minus losses, of course). Voltage times current in equals voltage times current out. This is akin to the rule of force times distance on one side of a lever equals force times distance on the other side.

The fundamental component of a transformer is an inductor. An inductor is simply a coil of wire, as we learned earlier. The number of turns of wire controls the concentration of the magnetic field. The core of the inductor also has the effect of concentrating the field. The material in the core can become saturated,[35] meaning that it cannot concentrate the field any more tightly than it has.

The important things to remember are that current creates a magnetic field, and a changing magnetic field creates a current. The changing field can be externally applied from a moving magnet, the input side of a transformer, or from the collapse of the field just created by the current. *Current* and *magnetic* fields are closely connected.

The Electric Field

Also called the *electrostatic field*, the electric field[36] is not as commonly known *per se* as the magnetic field. In the same way that current is connected to the magnetic field, voltage is connected to the electric field. That leads to a good rule of thumb to remember: *Current is magnetic* and *voltage is electric*.

The electric field comes from electric charges, both positive and negative. In a way that is analogous to the way like poles on magnets repel and opposite poles attract, like charges repel and opposite charges attract. Any molecule or atom can be neutral (no net charge), positively charged, or negatively charged.

---

[35] The topic of saturation is beyond the scope of this text, but it does occur and you can think of it like a sponge; in the same way that a sponge can soak up only so much water, the core of an inductor can help focus only so much magnetic field.

[36] I discuss the magnetic field first because it is better known to the average person. However, one can argue (as is pointed out in Chapter 0) that the electric field is more fundamental. The electric field happens because charge is there, whereas you get the magnetic field when you "move" the charge.

The accumulation of these charges is what is known as *voltage*. One way to think of it is that the *charges* are the *voltage* making the electric field, and *movement of those charges* is called *current* and creates the magnetic field.

Similarly to the way an inductor is a way of concentrating a magnetic field, a capacitor is a way of concentrating an electric field. Capacitors are made by two collectors or plates separated by a material that will not conduct electricity, also known as a *dielectric*. The symbol of a capacitor mimics the construction, shown in Figure 2.45.

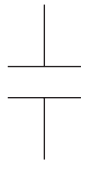


FIGURE 2.45
Capacitor symbol.

Because of the dielectric, current or actual charges cannot flow or move across the capacitor, all the charges build up on one side of the cap, kind of like a 50-car pileup on the freeway, as shown in Figure 2.46.

As the charges pile up on one side, the electrostatic field builds up, causing all the like charges on the other side of the cap to go rushing away (remember how like charges repel). Once it all comes to rest, there is an equal number of opposite charges on the other side of the cap. In this way the capacitor stores a charge of voltage on the plates of the capacitor.

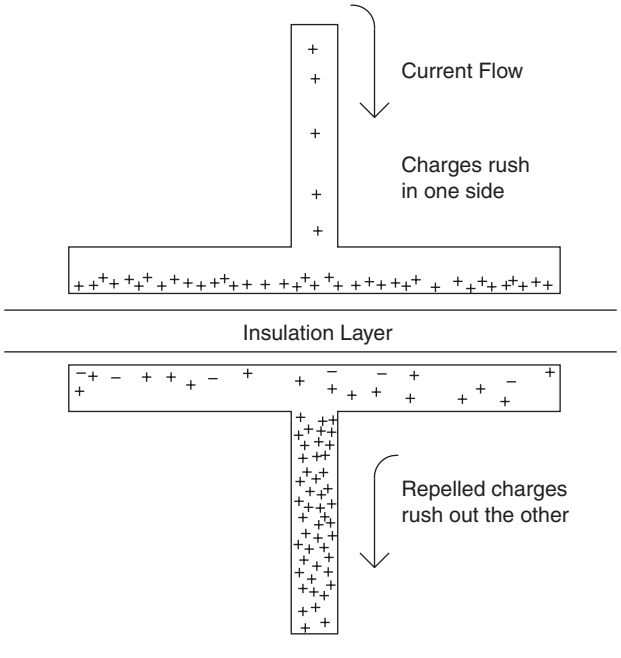


FIGURE 2.46
Behavior of charges in a capacitor.

How much charge a cap can store in an electric field is a function of the area of the plates. The amount of voltage it can store is dependent on the strength of the dielectric. If you exceed the capability of the insulation, the dielectric will break down and a charge will cross the gap. The same thing happens on a stormy day. During a thunderstorm charges build up in the clouds and the ground in the same way they do on either side of a capacitor. A lightning strike is a large-scale version of what happens when the insulation or dielectric in a capacitor breaks down.

In the same way current creates a magnetic field, voltage creates an electric field. Just as the magnetic field can store energy, the electric field can also store energy. As the magnetic field dissipates, it tries to maintain current. As the electric field dissipates, it tries to maintain voltage. Voltage and electric fields are closely connected.

Thumb Rules

- An inductor stores energy in a magnetic field.
- A capacitor stores energy in an electric field.
- Current is magnetic.
- Voltage is electric.

KEEP IT UNDER CONTROL

The topic of control theory is typically left till later in most education programs because it is considered a more advanced topic. Control systems, however, are a very common application in the electronic realm. Think about it; I'll bet most of the time you design a device to control something you don't want to think about, something that automatically does what it should without intervention. On top of that, control theory turns out to be useful in more than just clear-cut control applications—understanding op-amps and designing power supplies, for example. Since a basic knowledge of this concept will help you in many aspects of your career, it seems prudent to dedicate a few pages to it.

The System Concept

A system is anything with an input and an output. The idea is simple—take the input, shake it, squeeze it, do whatever, and then send it to the output. It can be represented with a block diagram, something like the one shown in Figure 2.47.

FIGURE 2.47
The magic box inside a system.

All the magic happens inside the box. This magic is called the *transfer function*. The transfer function is equal to the input over the output. It is the equation that you process the input through to get the output, so the following is true:

$$Output = Magic * Input \qquad \textbf{Eq. 2.17}$$

A little algebra yields:

$$\frac{Output}{Input} = Magic \qquad \textbf{Eq. 2.18}$$

Now you know how to find what the magic inside the box is, and sometimes it is just that easy. Let's try a simple example to see how. You put *12 miles* into the input, wait … chugga, chugga, ding! … and out pops *19.32 km*. As you might have guessed, the magic in this box is a metric converter, but what is the transfer function? According to the preceding equation, we simply divide the output by the input. That would be:

$$\frac{19.32\ km}{12\ miles} = 1.61\frac{km}{mile} \qquad \textbf{Eq. 2.19}$$

The magic in our converter box looks like the one in Figure 2.48.

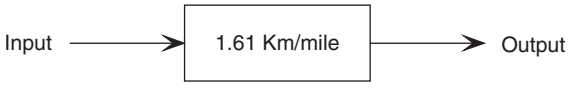


FIGURE 2.48
The magic revealed.

Note that the units made it into the box. This helps identify the type of units that will work at the input and what you will get at the output. Hopefully a little voice in your head is saying, "Isn't this a rehash of the unit math chapter?" It is, but this is a more formalized concept with some neat touches such as

the cool little boxes[37] you draw to help you understand the system. The next question you should ask is, "How does this apply to electronics?" Well, from the most basic to the most complex, you can represent any circuit with one of these magic (okay, some texts call them *black*) boxes.

We'd better do another example. Take a resistor. A resistor can be thought of as a current-to-voltage converter. Put current into the input, apply magic, and get voltage at the output. What would be the transfer function of that? If you mumbled a phrase with the words *Ohm's Law* anywhere in it, you are probably right.

$$R = \frac{V}{I} \qquad \textbf{Eq. 2.20}$$

That would make the block diagram of the resistor look something what Figure 2.49 shows.

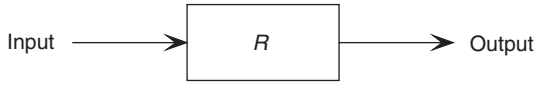


FIGURE 2.49
System diagram of a resistor.

In this transfer function, R is the value of the resistor in ohms, just in case you didn't guess. (Note that 1 unit of ohms equals 1 unit of volts divided by 1 unit of amps, like good old Ohm's Law says it does.)

Let's step up the idea to something a little more complex, like a voltage divider. We already know the equation for this. It is:

$$Vo = Vi \frac{Rg}{Rg + Ri} \qquad \textbf{Eq. 2.21}$$

Do you remember what *Vo* stands for? How about *Vi*? They are voltage output and voltage input. So let's just use a little algebra to figure this out. The transfer function is equal to the output over the input, like this:

$$\frac{Vo}{Vi} = \frac{Rg}{Rg + Ri} \qquad \textbf{Eq. 2.22}$$

[37]Drawing cool little boxes is fun! Why else do we represent what we think with schematics and flow charts and the like? I suggest that way back when we were all apes around a campfire, somehow drawing lines in the dirt gave us an evolutionary advantage. It must be important to the survival of the species, 'cause I will say this: I loved to get out the crayons as a kid, and a good time sketching on the white board is still pretty darn satisfying!

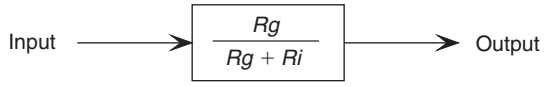


FIGURE 2.50
System diagram of a voltage divider.

The block diagram would look like Figure 2.50.

This same concept can be used to describe all the circuits that we have seen so far. You might see block diagrams of this type where *C* or *L* has a little *s* by it. This is a mathematical trick known as a *Laplace transform*. It is used to simplify problem solving. If you transform all that time constant and frequency stuff using Laplace pairs, you can treat the transformed equations with simple algebra and then transform it all back when you are done. Laplace transforms are beyond the scope of this text, but do take note that the *s* rolls up all the frequency response of capacitors and inductors into a domain that can be handled easily when the equations get complex.

We can describe any system as a magic box with an input and an output. One way to determine what is in the magic box is to put a known signal into the input. Let's take a look at what is probably the most important stimulus you can apply to the input of the magic box.

The Step Input

The idea behind the step input is to understand the output of a system by its response to a given input. The *step input* is an instantaneous change of state from a value of zero to another predetermined value. It looks like the one shown in Figure 2.51.

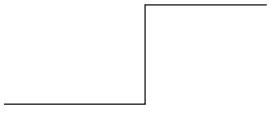


FIGURE 2.51
The step input.

The output will change in some manner predictable (one hopes!) by an equation. This is known as the *response*. The better you know the response of various components to this step input, the better you will be able to apply intuitive signal analysis. Let's go over the RC circuit again. The equation we learned earlier is:

$$Vo = Vi\left(1 - e^{-\frac{\tau}{rc}}\right) \qquad \text{Eq. 2.23}$$

To turn that into the transfer function for the magic box, all we have to do is move *Vi* to the other side of the equal sign, like this:

$$\frac{Vo}{Vi} = \left(1 - e^{-\frac{\tau}{rc}}\right) \qquad \text{Eq. 2.24}$$

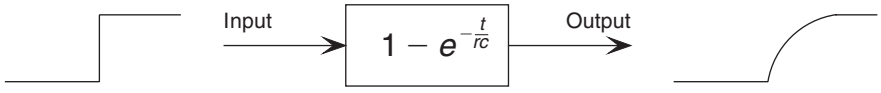


FIGURE 2.52
Step input into magic box with RC circuit inside.

We now plug that into the box. When we put the step input into one side, we get the familiar RC curve on the output, as shown in Figure 2.52.

Though it is nice to see the same conclusion from a different approach, where the system concept comes into its own is when we create a feedback loop.

Feedback

One of the neatest applications of control theory occurs when we implement feedback. *Feedback* is the process of using the output of the "magic box" as some portion of the input. Feedback comes in two flavors: positive and negative. They can be thought of in terms of interaction.

Positive Feedback

Positive feedback encourages or reinforces a behavior; *negative feedback* corrects or controls a behavior. For example, if your son is doing a good job in a soccer game and you cheer him on, which encourages him to try even harder, this is positive feedback. Positive feedback reinforces the behavior you desire. In the preceding case, it will encourage him to try as hard as possible. In fact, in a perfect world he will keep trying until he is giving it all he can. The same thing happens with positive feedback in control theory. Output is fed back to the positive input. This has the effect of increasing the output, which is fed back to the input, which will increase the output and so on (reinforcing the behavior) until the output is as high as it can go.

Since positive feedback reinforces the signal, the output can often "stick" at the rail.[38] For this reason the amount of positive feedback allowed is typically very limited, allowing only small changes to the input. These small changes can create a feature called *hysteresis*.

[38] If you don't understand this term, I refer you to the glossary, where I promise if you read the whole thing you get at least a few chuckles (so long as my editor doesn't take out all the jokes this time).

Another interesting thing that can happen due to this reinforcing behavior occurs when delays are created in the positive feedback loop. Think about it for a moment: What will happen if this signal is delayed a bit? If you time it right, the signal to change the output can be made to occur at the input when the output is already moving in the opposite direction. When this happens you have created an oscillator.

Now, though positive feedback is great for controlling toddlers, when it comes to circuits, if you want to control something you need negative feedback.

Negative Feedback

Negative feedback is a control situation. Let's go back to the soccer analogy for a moment. In this case, your son kicks the ball too far ahead of the player he is passing it to. You tell him to shorten up his pass. If he is not passing far enough, you tell him to lengthen it out. Based on how close the actual result is to the desired result, a corrective signal is fed back to the input. This corrective signal has a negative impact on the output, hence the term *negative feedback*.

Humans have an innate ability to handle negative feedback.[39] You probably experienced it this morning as you drove your car to work. If you drifted too close to the edge of the lane on the freeway, you processed a little negative feedback, resulting in a corrective signal to bring the car back to the center of the road. If you didn't, you are probably reading this in the passenger seat of the tow truck as your mangled car is hauled home!

Negative feedback is often used to create controlled amplifiers and filters. We will get into some details of negative and positive feedback and how they work using op-amps a bit later in the book.

Open-Loop Gain vs. Closed-Loop Gain

When you cut the feedback signals out of the "loop," the gain of the system is known as the *open-loop gain*. This is to distinguish it from the *closed-loop gain*, or the gain of the system when the feedback is in place. High open-loop gains in conjunction with negative feedback will minimize errors in amplifier and filter circuits.

---

[39]Unless, of course, it is coming from your significant other; then we tend to have serious systemwide erratic behavior.

Thumb Rules

- Lump everything into one magic box.
- The gain or magic equals the output over the input.
- Feedback loops can be added to these systems to create different results.
- Positive feedback tends to latch up or go to an output rail.
- Positive feedback delays can create oscillations.
- Negative feedback signals are corrective in nature.
- Negative feedback creates a controlled output.
- The gain of the system from input to output when the feedback is disconnected is known as the *open-loop gain.*
- The gain of the system when feedback is in place is known as the *closed-loop gain.*

CHAPTER 3

Pieces Parts

It takes parts to make a circuit, and lots of pieces, too. The better you know how these "pieces parts" work, as a friend of mine likes to say, the better stuff you will build.

PARTIALLY CONDUCTING ELECTRICITY

Semiconductors

Texts are available that can give you the quantum mechanical principles on which a semiconductor works. However, in this context I think the better thing to do is to give you a basic intuitive understanding of semiconductor components.

First, what is a semiconductor? *Conductor* in this case refers to the conduction of electricity. Think of a semiconductor as a material that partially conducts electricity or a material that is only semi-good at conducting electricity. It is similar to the resistor[1] that we just learned about; it's a component that will conduct electricity but not easily. In fact, the more you push through it, the hotter it gets as it resists this flow of electricity.

[1]Though the ubiquitous resistor originally was just really thin wires (you can still find that type in power resistors), these days in terms of sheer quantities most resistors are based on a semiconductor.

Before we move on, there is one other point to make. The world of semiconductor devices can be grouped into two categories: current driven and voltage driven.[2] Current-driven parts require current flow to get them to act. Voltage-driven devices respond to a change in voltage at the input. How much current or voltage is needed depends on the device you are dealing with.

Diodes

We will start our discussion with the diode. A *diode* is made of two types of semiconductors pushed together. They are known as type *P* and type *N*. They are created by a process called *doping*. In doping the silicon, an impurity is created in the crystal that affects the structure of the crystal. The type of impurity created can cause some very cool effects in silicon as it relates to electron flow.

Some dopants will create a type *N* structure in which there are some extra electrons simply hanging out with nowhere to go. Other dopants will create a type *P* structure in which there are missing electrons, also called *holes*. So we have one type, *N*, that will conduct negative charges with a little effort. We have another type that not only does not conduct but actually has holes that need filling. A cool thing happens when we smash these two types together; Figure 3.1 shows a sort of one-way electron valve known as the *diode*.

FIGURE 3.1
The PN junction of the diode.

Due to the interaction of the holes and the free electrons,[3] a diode allows current to flow in only one direction. A perfect diode would conduct electricity in one direction without any effect on the signal. In actuality, a diode has two

[2] I have seen texts argue this point as "current" really being simply movement of charges (that are "voltage"). However, I believe that using these categories will help you gain an intuitive understanding of these parts.

[3] When smashed together, some of the free electrons in the type *N* material crowd up next to the type *P* material (they are attracted there because of the positive charge). This creates what is known as a *depletion region*—an area where there aren't any free electrons (or charges) to move around, effectively blocking current flow. When you apply a voltage in the correct polarity on the diode, this region gets filled up with free charges, and thus current can pass through it.

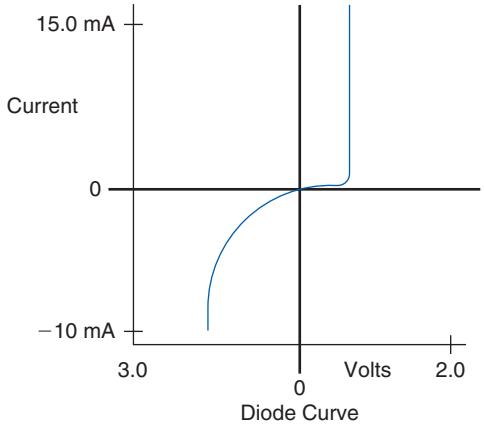


FIGURE 3.2
Typical diode voltage–current response.

important characteristics to consider: the forward voltage drop and the reverse breakdown voltage—see Figure 3.2.

Forward Voltage

The *forward voltage* is the amount of voltage needed to get current to flow across a diode. This is important to know because if you are trying to get a signal through a diode that is less than the forward voltage, you will be disappointed. Another often overlooked fact is the forward voltage times the current through the diode is the amount of power being dissipated at the diode junction. If this power exceeds the wattage rating of the diode, you will soon see the magic smoke come out and the diode will be toast.

For example, you have a diode with a forward voltage rating of 0.7 V and the circuit draws 2 A. This diode will be dissipating 1.4 W of energy as heat (just like a resistor). Verifying that your selection of diode can handle the power needed is an important rule of thumb.

Reverse Breakdown Voltage

Although a perfect diode could block any amount of voltage, the fact is, just like humans, every diode has its price. If the voltage in the reverse direction gets high enough, current will flow. The point at which this happens is called the *breakdown voltage* or the *peak inverse voltage*.[4] This voltage usually is pretty

[4]It is interesting to note that there is a type of diode called a *zener* in which this breakdown voltage is controlled and counted on. I would further stress the importance of calculating power in a zener. In this case, however, it is the zener voltage or the reverse voltage that you must multiply by current to calculate the power dissipation. Isn't *zener* a cool word to say?!

high, but keep in mind that it can be reached, especially if you are switching an inductor or motor in your circuit.

Transistors

The next type of semiconductor is made by tacking on another type *P* or type *N* junction to the diode structure. It is called a *BJT*, for *bipolar junction transistor*, or *transistor* for short. They come in two flavors: NPN and PNP; see Figure 3.3. I presume you can guess where those labels came from.

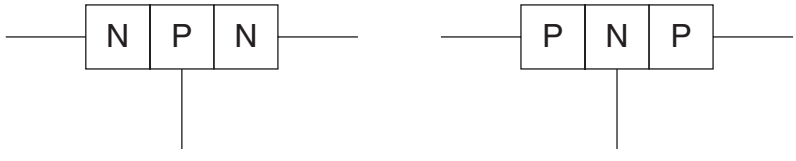


FIGURE 3.3
Smash diodes together to make a transistor.

At first glance you would probably say, "Isn't this just a couple of diodes hooked up back to back? Wouldn't that prevent current from flowing in either direction?" Well, you would be correct. It is a couple of diodes tied together, and yes, that prevents current flow. That is unless you apply a current to the middle part, also known as the *base* of the transistor. When a current is applied to the base, the junction is energized[5] and current flows through the transistor. The other connections on the transistor are called the *collector* and the *emitter*.

The NPN needs current to be pushed into the base to turn the transistor on, whereas the PNP needs current to be pulled out of the base to turn it on.[6] In other words the NPN needs the base to be more positive than the emitter, whereas the PNP needs the base to be more negative than the emitter. Remember the similarity to the diode? It is so close that the base-to-emitter junction behaves exactly like a diode, which means that you need to overcome the forward voltage drop to get it to conduct.

[5]Like the diode, charges from the base connection fill up the depletion region and thus current can begin flowing.

[6]In this case I am referring to *conventional flow*, as it is called. For more about this, read the AC/DC and a Dirty Little Secret section in Chapter 2.

Whoever is in charge of making up component symbols has made it easy for us. There is a very "diode-like" symbol on the emitter-to-base junction that indicates the presence of this diode. Also, please note that I keep talking about current into and out of the base of the transistors. Transistors are current-driven devices; they require significant current flow to operate. Most times the current flow needed in the base is 50 to 100 times less than the amount flowing through the emitter and collector, but it is significant compared to what are called *voltage-driven devices*.

Transistors can be used as amplifiers and switches. We should consider both types of applications.

Transistors as Switches

In today's digital world, transistors are often used as switches amplifying the output capability of a microcontroller for example. Since this is such a common application, we will discuss some design guidelines for using transistors in this manner.

Saturation

When you use a transistor as a switch, always consider if you are driving the device into saturation. *Saturation* occurs when you are putting enough current into the base to get the transistor to move the maximum amount through the collector. Many times I have seen an engineer scratching his head over a transistor that wasn't working right, only to find that there was not enough current going into the base.

Use the Right Transistor for the Job

Use an NPN to switch a ground leg and a PNP to switch a *Vcc* leg. This might seem odd to you at first. After all, they are both like a switch, right? Well, they are like a switch, but the diode drop in the base causes an important difference, especially when you only have 0 to 5 V to deal with. Consider the two designs shown in Figure 3.4 on the next page.

Let's do a little ISA[7] on the less robust circuit. As you decrease the voltage at the input, current will flow through the base, but the emitter base junction is a diode, right? That means that whatever voltage the base is at, the emitter is always

[7] Intuitive signal analysis—see Chapter 1. I have to get an acronym out there if I am to change the engineering world. Too bad all good acronyms mean more than one thing!

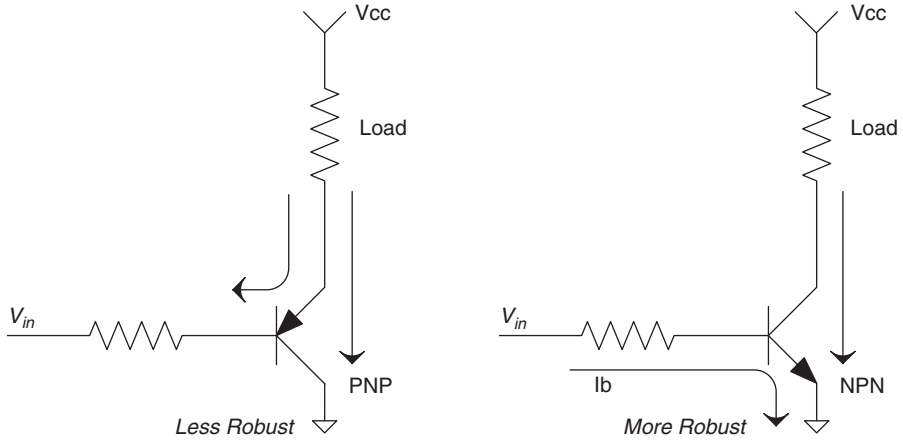


FIGURE 3.4
Comparison of different transistors in the same circuit.

0.7 V higher. Even if you get the input to 0 V exactly, since current has to flow, the voltage at the base will be a little higher. The voltage at the emitter will be 0.7 V above that. Notice now that any voltage change at this point will be reflected at the output. Now contrast that with the more robust design. When you pull the signal at the input low, current will flow through the base just like the other design, but do you see the difference? In the second design, the input voltage can vary quite a bit, and as long as the transistor is in saturation, the voltage drop at the output from collector to emitter will remain the same—see Figure 3.5.

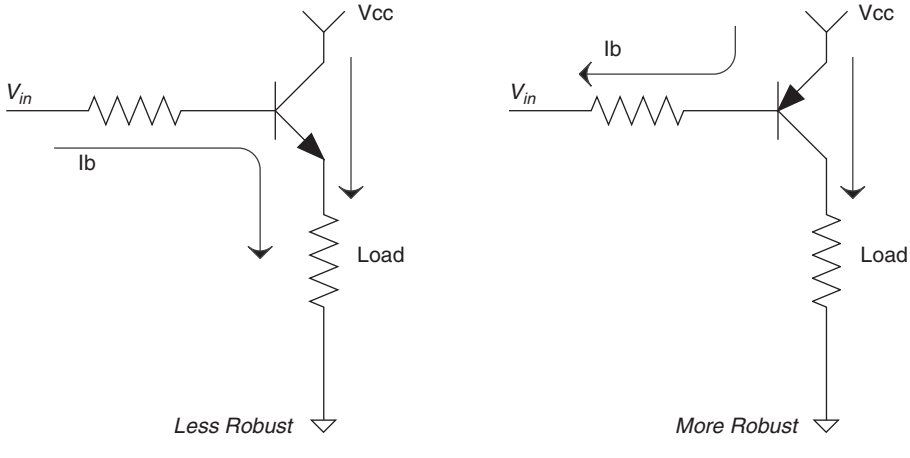


FIGURE 3.5
Comparison of different transistors in the same circuit.

The PNP transistor works best in the opposite configuration. For a switching application it is more robust when it controls the *Vcc* leg of the load. In both cases turning the transistor off is not too difficult; just get the base within 0.7 V of the emitter and the current will stop flowing.

Transistors as Linear Amplifiers

Transistors can also be used as linear amplifiers. This is because the amount of current flowing through the collector is proportional[8] to the current through the base. This is called the *beta* or *HFE* of the transistor. For example, if you put 5 μA into the base of a transistor with a beta of 100, you would get 0.5 mA of collector current. Making this work correctly depends on keeping the transistor operating inside a couple of important limits.

One limit is created by the diode in the base-to-emitter connection. This diode needs to remain forward-biased for the transistor to amplify linearly. It is also important to keep the transistor out of saturation. This can push the transistor out of its linear region, creating funny results such as clipping. What all this means is that setting up linear transistor amplifiers can be a bit of a trick. You need to pay attention to biasing and the HFE, which unfortunately varies considerably from part to part. These days I rarely use transistors alone as linear amplifiers for two reasons: the first is the amount of variation from part to part mentioned before (a real issue when you make millions of circuits), and the second is the fact that operational amplifiers (which we will discuss later) are so inexpensive[9] and easy to use. If you need the power capability of a transistor, you should try teaming it up with an op-amp to make life easier!

FETs

FETs, or *field effect transistors*, were developed more recently than transistors and diodes. Why come up with something new? Simple: FETs have some properties that make them very desirable components. The primary reason they are so slick is that the output of a FET is basically a resistance that varies depending on the voltage at the input. The outputs on a FET are called the *drain* and *source*, whereas the input is known as the *gate*.

Virtually no current is needed at the gate to affect an FET; this makes it an ideal component for amplifying a signal that is weak, since the FET will not load the

[8]This is also the reason that they are often referred to as *current-driven* devices.

[9]You can buy a quad op-amp for less than three or four transistors these days, so why make it hard on yourself if you don't have to?

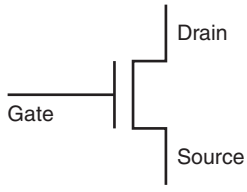


FIGURE 3.6
The FET.

signal significantly. In fact, some of the better op-amps use FETs at their inputs for just this reason. One downside to an FET is that the parts tend to be easier to break than their transistor cousins. They are sensitive to static and over-voltage conditions, so be sure to pay attention to the maximum ratings when you use these parts.

One very cool thing about an FET is the drain-to-source connection. It acts just like a resistor that you control by the voltage at the gate. This in effect makes it an electronically controlled variable resistor. For this reason, it is common to find FETs in circuits creating variable gain control. The drain-to-source connection acts like a resistor in either direction. That is, current can flow either way. However, you should expect a FET to have a built-in reverse-biased diode across the drain-to-source pins. (It is the nature of the construction of the FET that creates this diode.)

When used in switch mode, a term you should pay attention to is *RDSon*. This is the *resistance drain to source* when the device is turned all the way on. The lower this number, the less power you will lose across the device as heat. The voltage across the device will be the current times RDSon, and the power dissipated in heat will be this voltage times the current through the device.

An ohm equals volts divided by current if Ohm's Law still holds true (by this point in the book, a resounding *Yes!* should be on the tip of your tongue). The inverse of an ohm or $1/R$ equals current divided by voltage. This is known as a *mho*.[10] Mhos are to FETs as beta or HFE is to a transistor. This is the unit of gain, also known as *transconductance*, that defines the output of the FET. Put X volts into the gate of the FET, times that by the transconductance, and you will get Y current drain to source.

[10]This unit is also known as a *Siemens*, after that well-known brand name on many electronic gadgets you see around today. (Okay, so it is really named after the guy who started the company that makes the stuff today.) Anyway, I like mho better; it just makes sense, since it is the inverse of an ohm after all. I still have no idea as to the origin of the word mho. Drop me a line if you know where it came from!

Just as with transistors, this gain from input to output varies significantly from part to part. When using the transistors in linear mode, you need to either characterize the component you are using or develop some type of feedback control method that compensates for the variation to achieve the desired result.

In my experience, some engineers really like FETs and some like the good old BJT. I say keep both tools in your tool chest and use the right one for the job at hand.

Random List of Additional Parts

Here are a few parts in the semiconductor world that you might or might not have heard of:

Darlington transistor. This type of transistor consists of two transistors hooked together to increase the gain, as can be seen by the symbol used to represent it. Note that the base emitter diode drop is basically doubled in a Darlington transistor.

SCR. This is what you get when you create a PNPN junction, called a *silicon-controlled rectifier*. Basically the combination of a diode and a transistor, it can switch large currents easily. But one caveat—you can turn it on but not turn it off. The current through the SCR must get below the holding current (very small) before it turns itself off. The SCR is part of the thyristor family.

TRIAC. This is a cousin to the SCR and also is in the thyristor family. Think of it as two SCRs back to back, making it an effective AC switch. It is often found in solid-state relays and the like.

IGBT. The *isolated gate bipolar transistor* is best thought of as a combination between a transistor and a FET. An FET is used to push a load of current through a big transistor.

There aren't really a lot of different variations in semiconductors; they all boil down to some basic configurations of the *P* and *N* materials. It is amazing to me that such a level of complexity is achieved from just a few parts, but semiconductors have truly revolutionized the world as we know it today. The devil is in the details, however. I can't stress too much the need to look at the datasheet of the part you are using. The more you know about its idiosyncrasies, the better your designs will be.

Thumb Rules

- Diodes are a "one-way" valve for electrons.
- Diodes have a forward-voltage drop you must overcome before they will conduct.
- Transistors are current driven.
- Transistors have a diode in the base that needs to be biased to work right.
- When using transistors as switches, check saturation current.
- FETs are voltage driven.
- FETs tend to be less robust; take care to design plenty of headroom between your circuit and the maximum ratings of the part.
- FETs are static sensitive.
- Meticulously study the datasheet of the part you are using.

POWER AND HEAT MANAGEMENT

One thing in common with all electrical devices (this side of superconductors) is the fact that as they operate, heat is generated. This is because in every component (as we will learn later) there is some amount of equivalent resistance. Resistance times current flow equals a voltage drop, and a voltage drop times current equals power. Since Ohm's Law is unavoidable, this power must turn into heat. Heat is the premier cause of wear and tear in electronic components, so managing heat is a good thing to know something about. Let's start from the inside out.

Junction Temp

Inside a semiconductor, the place where all the magic happens is called the *junction*. This is the point where all the heat comes from as the part operates. The junction will have a maximum temperature that it can reach before something goes wrong. You guessed it; you find out just how much it can handle by reading the datasheet for the part.

Case Temp

The junction is always inside some type of case. Since you can't measure junction temperature when you need to test a design, you have to measure case temperature. There will always be a temperature drop from the junction to the

case. The amount will typically be indicated in the part's spec sheet. If it says the case-to-junction thermal drop is 15°C, expect the junction temp to be 15° warmer than what you measure. Here is where a good engineer will fudge the numbers in his favor. If the boss asks you to run this part as close to the edge as possible, tell her you need to be 30° under the junction temp per the spec sheet. Most likely she won't know where to look for this information, so will probably believe you and you will have a more robust design.

Heat Sinking

How hot the case gets depends on the heat sink attached to it. The case itself will be able to radiate a certain amount into the air around it. If this isn't sufficient, a *heat sink* can be added. One point you should recognize is that a heat sink (contrary to what you might think, given the name) is not a hole into which you can dump the heat from the part. A heat sink is more accurately described as a way to more efficiently transfer heat into the surrounding environment (this happens to be the air in most cases).

Heat sinks capture that thermal rise and dissipate it into the surrounding air. Heat sinks are rated by a °C/W number. This number represents how much the temperature of the device on the sink will rise for every watt of heat generated. For example, if you put 20 watts of heat on a 3°C/W heat sink, the power device hooked up to that heat sink will rise 60°C above the ambient temperature.

Heat sinks can be thought of as heat conductors. Just as some metals are better electric conductors than others, some metals are better heat conductors. Usually one goes with the other. Aluminum is a better electrical conductor than steel, and it is also a better heat conductor. Copper, one of the best electrical conductors around, is also one of the best heat conductors. Thought of in these terms, the heat sink conducts heat away from the part. Like the fact that current always flows in one direction, heat always flows from hot to cold. There are a couple of ways for this to happen, as we will see now.

Radiation

Once the heat sink is warm, it will emit infrared radiation; as this energy is radiated away, the heat sink will cool. Have you ever wondered why so many heat sinks are black? This is because the color black[11] is an efficient radiator, just as

[11]The color is not a major player when it comes to getting rid of heat, but it does help, so if you really need that last little bit of power handling, go black (but a little more metal will work just as well).

this color tends to absorb more infrared radiation (as you probably have noticed if you have ever worn a black shirt on a sunny day). It will radiate this heat away as well, as long as the part is in a cooler environment and the sun isn't shining on it! Although radiation is a way of getting heat moving away from your part, in most electronic devices today there are much better ways to get rid of heat.

Convection

The best way to get rid of heat is by moving some air across your heat sink. This is called *convection*. There are two ways to achieve convection: one is by placing the sink so that air that is warmed by proximity to the heat sink rises. As this happens, cooler air takes its place to be warmed up and the whole process repeats. (See Figure 3.7.) Most heat sinks have some type of spec as to free air operation that describes their function in this case.

One quick side note: Free air convection relies on the presence of gravity (hot air won't rise to be replaced by the cooler air without gravity), so if you happen to be working on a space shuttle experiment, don't count on free air convection for cooling!

A huge difference in cooling a heat sink can be achieved by moving more air across it. This is commonly accomplished by some type of fan. It is not unusual to see a heat sink handle 10 times as much power just by placing a fan next to it. This is the reason that so many devices these days have acquired that proverbial hum of a fan that is so prevalent.

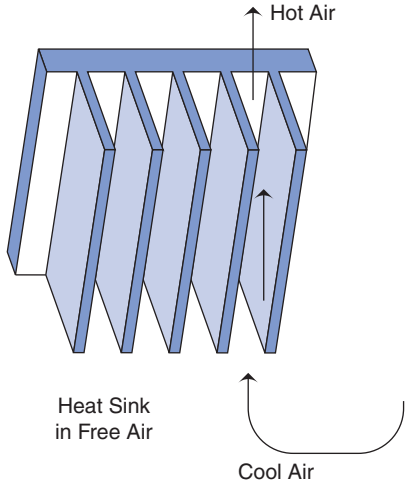


FIGURE 3.7
Convection on a heat sink.

The more heat sink area you have in contact with the air, the better it can transfer heat. For this reason, you will see a lot of fins on these parts. More fins mean more surface area, which means more efficient heat transfer.

Hmmm, here's a thought: Wouldn't it really be nice to recapture this heat and turn it back into power? I know there are thermoelectric devices that generate electricity when you heat them up, so this seems like a no-brainer. I guess I will get to that design later, but if any of you reading this get to the punch before me and make millions with this idea, all I ask is 1%!

Conduction

Another way of moving heat is by *conduction*. This is how the heat gets from the part into the heat sink, and it is how the heat travels across the sink as well. Conduction moves heat very, very well (that is how it gets from the part into the heat sink), but whatever it is conducting to must be cooler than where the heat is coming from in order for the heat to flow. Often a liquid is used to conduct heat away from stuff that gets hot, such as a nuclear reactor or your car engine. At the end of the day, though, that heat has to go somewhere. That's why you see a radiator in the front of your car dumping all that heat collected by the antifreeze into the atmosphere. The engine in my boat uses the entire lake as a heat sink, with no radiator needed, since it should be fairly obvious that my piddling little boat isn't going to have enough power to raise the average temperature of millions of gallons of water by even a fraction of a degree.[12]

Can You Dump It into a PCB?

This is a question that I have often heard: Can you use the PCB as a heat sink? The answer is yes. In fact, the PCB is simply copper plating, and we know that copper is a good heat conductor, so it follows that it can be used as a heat sink. Okay, here it comes . . . *but* . . . how do you know how well the PCB radiates the heat into the atmosphere? That is something you will most likely have to test to figure out. There are just so many variables in calculating this that it is

[12]You might even say, "Forget about the greenhouse effect—what about all this energy we are pouring into the atmosphere off our heat sinks?" If you consider that on average, every house in the world dumps 500 W of heat from light bulbs alone into the atmosphere, and you figure there are about a billion houses, that comes out to a lot of energy! Is it enough to raise the temperature of the Earth? I would have to dig a lot further back into chemistry classes than I would like to figure that out. However, since it is fun to simply spout generalities, I predict that sooner or later, if we keep making more heat, we will cook ourselves! Of course, if the sun were to sneeze even just a bit, we could find ourselves wishing we had those heaters going!

faster to lay out the PCB, stick the part on, and try it. Here are some items to note when you're using a PCB as a heat sink:

- A lot of little *via*s connecting the top layer to the bottom one will help increase the amount of surface area you have to dissipate the heat.
- The PCB in this area is going to get warm. That means expansion and contraction of the PCB. You might find that this could cause mechanical damage over time or even crack solder joints and PCB connections.
- I would recommend keeping the PCB heat sinks under 60°C. A cool rule of thumb I have learned is that if a metal surface is hot enough to burn you at the touch, it is more than 60°C.[13]

Heat Spreading

One of the major factors that control heat conduction when you have two materials next to each other is the surface area of the two materials that are touching. One other thing that affects conduction of a single material is the thickness of the material.

This gives rise to a technique known as *heat spreading*. A big, thick, very thermally conductive material is bolted up to the "hot part" to serve as a high-speed conduit to a bigger heat sink, where all the fins for radiating the heat are located.[14] The idea is to keep the junction temperature of the device lower by getting the heat away faster.

Does it work, you say? Truth is, it can work, but there are many variables involved (such as the thermal conductivity between the heat spreader block and the rest of the heat sink, for example). As in the case of using the PCB as a heat sink, you should take it to the test lab to see if it is really working well or even helping. Remember, though, there will be a temperature gradient everywhere that there is a junction between two parts; the fewer junctions, the better your heat sink will work.

---

[13] By no means am I endorsing touching a hot component as a way of checking its temperature! I hope that this disclaimer is enough to keep the lawsuit-happy people out there off my case. I wouldn't want anyone to get burned. I could go on about the legal ills that are crippling our world, but that is a whole other topic. Suffice it to say, if you happen to get burned by accident, you can be reasonably sure the metal you touched was more than 60°C. Please don't touch it on purpose; there are much more accurate ways of measuring temperature than by using your finger.

[14] If you take a close look at power heat sinks, you will notice a varying thickness in the aluminum, from the attachment point to the fins, that serves this very purpose.

Thumb Rules

- Meticulously study the datasheet of the part you are using (repeated for emphasis).
- Heat is the biggest killer of electronic components.
- Most heat sinks dump heat into the air around them, most commonly by convection.
- If a part burns you when you touch it, it is more than 60°C.
- You can use a PCB as a heat sink, but take care to test it.

THE MAGICAL MYSTERIOUS OP-AMP

Op-Amps: The Misunderstood Magical Tool!

In my opinion, op-amps are probably the most misunderstood yet potentially useful IC at the engineer's disposal. It makes sense that if you can understand this device, you can put it to use, giving you a great advantage in designing successful products.

What Is an Op-Amp, Really?

Do you understand how an op-amp works? Would you believe that op-amps were designed to make it *easier* to create a circuit? You probably didn't think that the last time you were puzzling over a misbehaving breadboard in the lab.

In today's digital world it seems to be common practice to breeze over the topic of op-amps, giving the student a dusting of commonly used formulas without really explaining the purpose or theory behind them. Then the first time a new engineer designs an op-amp circuit, the result is utter confusion when the circuit doesn't work as expected. This discussion is intended to give some insight into the guts of an operational amplifier and to give the reader an intuitive understanding of op-amps.

One last point: Make sure that you read this section first! It is my opinion that one of the causes of op-fusion (op-amp confusion), as I like to call it, is that the theory is taught out of order. There is a very specific order to learning the theory, so please understand each section before moving on. First, let's take a look at the symbol of an op-amp (see Figure 3.8 on next page).

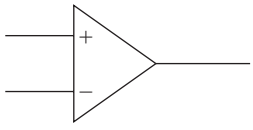


FIGURE 3.8
Your basic op-amp.

There are two inputs, one positive and one negative, identified by the + and − signs. There is one output.

The inputs are high impedance. I repeat. The inputs are high impedance. Let me say that one more time. *The inputs are high impedance!* This means that they have (virtually) no effect on the circuit to which they are attached. Write this down because it is very important. We will talk about this in more detail later. This important fact is commonly forgotten and contributes to the confusion I mentioned earlier.

The output is low impedance. For most analysis it is best to consider it a voltage source. Now let's represent the op-amp, as in Figure 3.9, with two separate symbols.

You see here a summing block and an amplification block. You may remember similar symbols from your control theory class. Actually, they are not just similar—they are exactly the same. Control theory works for op-amps. (There will be more on this topic coming up later.)

First, let's discuss the summing block. You will notice that there is a positive input and a negative input on the summing block, just as on the op-amp. Recognize that the negative input is as though the voltage at that point is multiplied by −1. Thus, if you have 1 V at the positive input and 2 V at the negative input, the output of this block is −1. The output of this block is the sum of the two inputs where one of the inputs is multiplied by −1. It can also be thought of as the difference of the two inputs and represented by this equation:

$$Vs = (V+) - (V-) \qquad \textbf{Eq. 3.1}$$

Now we come to the amplification block. The variable *G* inside this block represents the amount of amplification that the op-amp applies to the sum of

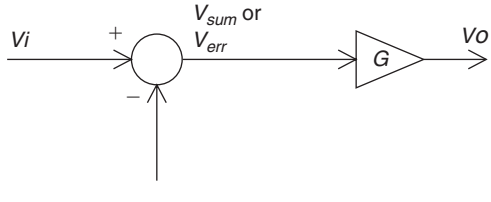


FIGURE 3.9
What is really inside an op-amp?

the input voltages. This is also known as the *open-loop* gain of the op-amp. In this case, we will use a value of 50,000. I hear you say, "How can that be? The amplification circuit I just built with an op-amp doesn't go that high!" Just trust me for a moment. We will get to the amplification applications shortly. Just go find the open-loop gain in the manufacturer's datasheet. You will see this level of gain or even higher is typical of most op-amps.

Now let's do a little analysis. What will happen at the output if you put 2V on the positive input and 3V on the negative input? I recommend that you actually try this on a breadboard. I want you to see that an op-amp can and will operate with different voltages at the inputs. However, a little math and some common sense will also show us what will happen. For example:

$$V_{out} = 50,000 * (2 - 3), \text{or} -50,000\,V \qquad \textbf{Eq. 3.2}$$

Now, unless you have a 50,000V op-amp hooked up to a 50,000V bipolar supply, you won't see −50,000V at the output. What will you see? Think about it a minute before you read on. The output will go to the minimum rail. In other words, it will try to go as negative as possible. This makes a lot of sense if you think about it like this. The output wants to go to −50,000V and obey the preceding mathematics. It can't get there, so it will go as close as possible. The rails of an op-amp are like the rails of a train track; a train will stay within its rails if at all possible. Similarly, if an op-amp is forced outside its rails, disaster occurs and the proverbial magic smoke will be let out of the chip. The rail is the maximum and minimum voltage the op-amp can output. As you can intuit, this depends on the power supply and the output specifics of the op-amp. Okay, reverse the inputs. Now the following is true:

$$V_{out} = 50,000 * (3 - 2), \text{ or} +50,000\,V \qquad \textbf{Eq. 3.3}$$

What will happen now? The output will go to the maximum rail. How do you know where the output rails of the op-amp are? As noted before, that depends on the power supply you are using and the specific op-amp. You will need to check the manufacturer's datasheet for that information. Let's assume that we are using an LM324, with a +5V single-sided supply. In this case the output would get very close to 0V when trying to go negative and around 4V when trying to go positive.

At this time I would like to point something out. The inputs of the op-amp are *not equal* to each other. Many times I have seen engineers expect these inputs to be the same value. During the analysis stage, the designer comes up with currents going into the inputs of the device to make this happen (remember, high

impedance inputs, virtually zero current flow). Then when he tries it out, he is confused by the fact that he can measure different voltages at the inputs.

In a special case we will discuss in the next section, you can make the assumption that these inputs are equal. *It is not the general case!* This is a common misconception. You must not fall into this trap or you will not understand op-amps at all.

The previous examples indicate a very neat application of op-amps: the comparator circuit. This is a great little circuit to convert from the analog world to the digital one. Using this circuit you can determine whether one input signal is higher or lower than another. In fact, many microcontrollers use a comparator circuit in analog-to-digital conversion processes. Comparator circuits are in use all around us. How do you think the streetlight knows when it is dark enough to turn on? It uses a comparator circuit hooked up to a light sensor. How does a traffic light know when there is car present above the sensors to trigger a cycle to green? You can bet there is a comparator circuit in there.

Thumb Rules

- The inputs are high impedance; they have negligible effects on the circuit to which they are hooked.
- The inputs can have different voltages applied to them; they do not have to be equal.
- The open-loop gain of an op-amp is very high.
- Due to the high open-loop gain and the output limitations of the op-amp, if one input is higher than the other, the output will "rail" to its maximum or minimum value. (This application is often called a comparator circuit.)

NEGATIVE FEEDBACK

If you didn't just finish reading them, go back and read the last section's thumb rules. They are very important in developing a correct understanding of what an op-amp does. Why are these points important? Let's go over a little history.

Up until the invention of op-amps, engineers were limited to the use of transistors in amplification circuits. The problem with transistors is that, being "current-driven" devices, they always affect the signal of the circuit that the designer

wants to amplify by loading the circuit. Also, due to manufacturing tolerances of transistors, the gain of the circuits would vary significantly. All in all, designing an amplifier circuit was a tedious process that required much trial and error. What engineers wanted was a simple device that they could attach to a signal that could multiply the value by any desired amount. The device should be easy to use and require very few external components. To paraphrase, *operation* of this *amplifier* should be a "piece of cake." At least that is the way I remember it. The other way the name *operational amplifier*, or *op-amp*, came into being was to describe the fact that these amplifiers were used to create circuits in analog computers, performing such *operations* as multiplication, among others.

To begin with, let's take a look at the special case I mentioned in the previous discussion. First, return to the previous block diagram and add a feedback loop, as shown in Figure 3.10.

You will see that I have represented the forward or open-loop gain with the value *G* and the feedback gain with the value *H*. The first thing you should notice is that the output is tied to the negative input. This is called *negative feedback*. What good is negative feedback? Let's try an experiment. Hold your hand an inch over your desk and keep it there. You are experiencing negative feedback right now. You are observing via sight and feel the distance from your hand to the desk. If your hand moves, you respond with a movement in the opposite direction. This is negative feedback. You invert the signal you receive via your senses and send it back to your arm. The same thing occurs when negative feedback is applied to an op-amp. The output signal is sent back to the negative input. A signal change in one direction at the output causes a V_{sum} to change in the opposite direction.

You should get an intuitive grasp of this negative feedback configuration. Look at the previous diagram and assume a value of 50,000 for *G* and a value of 1 for *H*. Now start by applying a 1 to the positive input. Assume that the negative

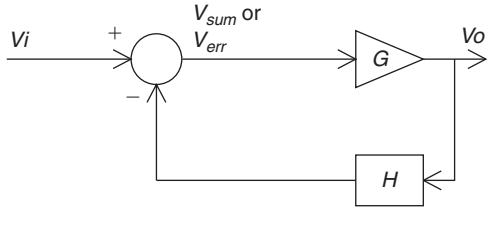


FIGURE 3.10
Adding negative feedback to the op-amp.

input is at 0 to begin with. That puts a value of 1 at the input of the gain block *G* and the output will start heading for the positive rail. But what happens as the output approaches 1? The negative input also approaches 1. The output of the summing block is getting smaller and smaller. If the negative input goes higher than 1, the input to the gain block *G* will go negative as well, forcing the output to go in the negative direction. Of course, that will cause a positive error to appear at the input of the gain block *G*, starting the whole process over again. Where will this all stop? It will stop when the negative input is equal to the positive input. In this case, since *H* is 1, the output will also be 1.

You have learned (or will learn) this in control theory. Look at the basic control equation in reference to the previous diagram:

$$Vo = Vi * \frac{G}{1 + G * H} \qquad \textbf{Eq. 3.4}$$

What happens when *G* is very large?[15] The 1 in the denominator becomes insignificant and the equation becomes:

$$Vo = \text{approximately } Vi * (1/H) \qquad \textbf{Eq. 3.5}$$

H in this case is 1,[16] so it follows that:

$$Vo = \text{approximately } Vi * (1/1) \qquad \textbf{Eq. 3.6}$$

or:

$$Vo = Vi \qquad \textbf{Eq. 3.7}$$

This is the special case in which you can assume that the inputs of the op-amp are equal. Apply it *only* when there is negative feedback. When feedback gain is 1, this also demonstrates another neat op-amp circuit: the voltage follower. Whatever voltage is put on the positive input will appear at the output.

Take a look at Figure 3.11. This is an op-amp in the negative feedback configuration. When you look at this, you should see a summer and an amplifier, just as in the previous drawing. In this configuration, you can make the assumption that the positive and negative inputs are equal.

Negative feedback is the case that is drilled into you in school and is the one that often causes confusion. It is a special case—a very widely used special case.

---

[15] Remember, an op-amp has a very large *G!*

[16] *H* doesn't have to be 1 for this special case to occur; there simply needs to be negative feedback present.

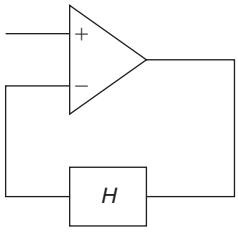


FIGURE 3.11
Original op-amp symbol with negative feedback.

Nonetheless, if you do not have negative feedback and the inputs and output are within operational limits, you must *not* assume that the inputs of the op-amp are equal.

Why is this negative feedback configuration used so much? Remember the reason that op-amps were invented? Amplifiers were tough to make. There had to be an easier way. Take a look at the control equation again:

$$Vo = Vi * \frac{G}{1 + G * H} \qquad \textbf{Eq. 3.8}$$

I have already shown that for large values of G, the equation approximates:

$$Vo = Vi * \frac{1}{H} \qquad \textbf{Eq. 3.9}$$

You will see that the amplification of Vi depends on the value of H. For example, if we can make H equal 1/10, then it follows that:

$$Vo = Vi * (1/(1/10)) \qquad \textbf{Eq. 3.10}$$

or:

$$Vo = Vi * 10 \qquad \textbf{Eq. 3.11}$$

How do we go about doing that? Do you remember the voltage divider circuit? That would be very useful here, since we would like H to be the equivalent of dividing by 10. Let's insert the voltage divider circuit in place of H.

Notice that the input to the voltage divider comes from the output of the op-amp Vo. The output of the voltage divider goes to the negative input of the op-amp $V-$. Now, will the op-amp input $V-$ affect the voltage divider circuit? *No!* It has high impedance. It will not affect the divider. (If you didn't get that, go back and read the "What Is an op-amp, Really?" section till you do!)

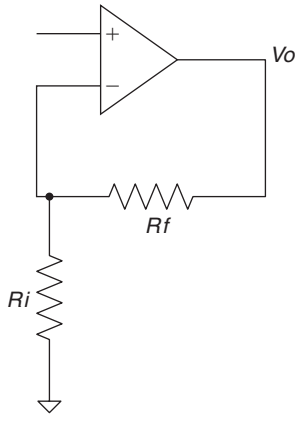


FIGURE 3.12
Negative feedback is a voltage divider.

Since the input to the divider is hooked to a voltage source, and the output is not affect by the circuit, we can calculate the gain from *Vo* to *V*− very easily with the voltage divider rule shown in Figure 3.12.

$$\frac{V-}{Vo} = \frac{Ri}{Ri + Rf} = H \qquad \textbf{Eq. 3.12}$$

Thus it follows that:

$$\frac{1}{H} = \frac{Ri + Rf}{Ri} \qquad \textbf{Eq. 3.13}$$

or, with a little algebra:

$$\frac{1}{H} = \frac{Ri}{Ri} + \frac{Rf}{Ri} = \frac{Rf}{Ri} + 1 \text{ or } \frac{1}{H} = \frac{Rf}{Ri} + 1 \qquad \textbf{Eq. 3.14}$$

There you have it—the gain of this op-amp circuit. Let's look at it another way. Go back to the previous equation:

$$\frac{V-}{Vo} = \frac{Ri}{Ri + Rf} \qquad \textbf{Eq. 3.15}$$

We learned that in this special case of negative feedback, we can assume that *V*+ =*V*−. This is because the negative feedback loop is pushing the output around, trying to reach this state. So let's assume that *Vi* = *V*+ , which is where the input to our amplifier will be hooked up. Now we can replace *V*− with *Vi*, and the equation looks like the following:

$$\frac{Vi}{Vo} = \frac{Ri}{Ri + Rf} \qquad \text{Eq. 3.16}$$

What we really want to know is, what does the circuit do to *Vi* to get *Vo*? Let's do a little math to come up with this equation:

$$Vo = Vi * \frac{Ri + Rf}{Ri} = Vi * \frac{Rf}{Ri + 1} \text{ or } \frac{Vo}{Vi} = \frac{Rf}{Ri} + 1 \qquad \text{Eq. 3.17}$$

Please note that this is equal to $1/H$. You see, the gain of this circuit is controlled by two simple resistors. Believe me, this is a whole lot easier to define and calculate than a transistor amplification circuit. As you can see, the operation of this amplifier is pretty easy to understand.

Thumb Rules

- The negative feedback configuration is the only time you can assume that $V- = V+$.
- The high impedance inputs and the low impedance output make it easy to calculate the effects simple resistor networks can have in a feedback loop.
- The high open-loop gain of the op-amp is what makes the output gain of this special case equal to approximately $1/H$.
- Op-amps were meant to make amplification easy, so don't make it hard!

POSITIVE FEEDBACK

What is positive feedback? Let's take a look at a real-world example. You are hard at work one day when your boss stops by and says, "Hey, you should know that you've handled your project very well, and that new op-amp circuit you built is awesome!" After you bask in his praise for a while, you find yourself working even harder than before.* This is *positive feedback*. The output is sent back to the positive input, which in turn causes the output to move further in the same direction. Let's look at the op-amp diagram again—see Figure 3.13 on the next page.

Now we will do a little intuitive analysis. Don't forget the Thumb Rules we learned in the last two sections. Review them now if you need to.

*OK, this is only true if you actually believe your boss.

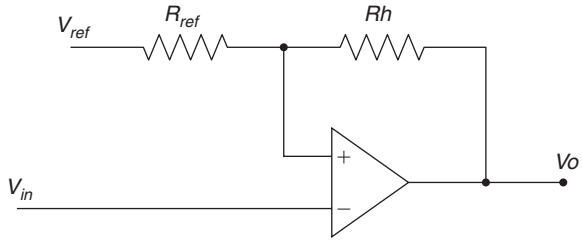


FIGURE 3.13
Positive feedback on an op-amp.

Begin by applying 0 V to V_{in}. In this case the input is connected to $V-$. You also see that the output is connected via a resistor to a reference voltage, V_{ref}. What is the voltage at $V+$? Does the voltage at $V+$ equal the voltage at $V-$? *No!* (Don't believe me? Check the Thumb Rules!)

What *is* the voltage at $V+$? That depends on two things: the voltage at V_{ref} and the output voltage of the amplifier, *Vo*. Does the $V+$ input load the circuit at all? No, it does not. To begin the analysis, let $V_{ref} = 2.5$ V, and assume that the output is equal to 0 V. Now what is the voltage at $V+$? What do you know—since *Vo* is equal to 0, we have a basic voltage divider again. Assume $R_{ref} = 10$ K and $Rh = 100$ K:

$$V+ = V_{ref} * \frac{Rh}{Rh + R_{ref}} = 2.5 * \frac{100\,\text{K}}{110\,\text{K}} = 2.275\,\text{V} \qquad \textbf{Eq. 3.18}$$

So now there is 2.275 V at $V+$ and 0 V at $V-$. What will the op-amp do? Let's refer to the op-amp block diagram we learned earlier—see Figure 3.14.

What do we have? V_{sum} is equal to $V+ \; - \; V-$ or, in this case, $V_{sum} = 2.275$ V. *Vo* is equal to $V_{sum} * G$. The output will obviously go to the positive rail. (If this is not obvious to you, you need to review "What Is an Op-amp, Really?" again.) Now we have *Vo* at the positive rail. Let's assume that it is 4 V for this particular op-amp. (Remember, the output rails depend on the op-amp used, and you should always refer to the datasheets for that information. 4 V used in this case is typical for an LM324 with a 0 to 5 V supply.)

The output is at 4 V and $V-$ is at 0 V, but what about $V+$? It has changed. We must go back and analyze it again. (Do you feel like you are going in circles? You should. That is what feedback is all about; outputs affect inputs, which affect the outputs, and so on, and so on.) The analysis this time has changed slightly. It is no longer possible to use just the voltage divider rule to calculate $V+$. We must also use *superposition*.

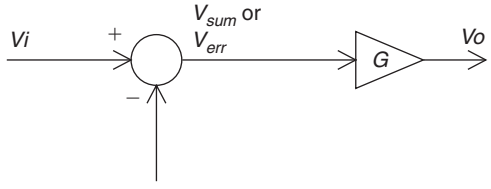


FIGURE 3.14
Start with what is really inside.

In superposition, you set one voltage source to 0 and analyze the results, and then you set the other source to 0 and analyze the results. Then you add the two results together to get the complete equation. Let's do that now. We already know the result due to V_{ref} from our previous example. Figure 3.15 shows the positive feedback diagram again for reference.

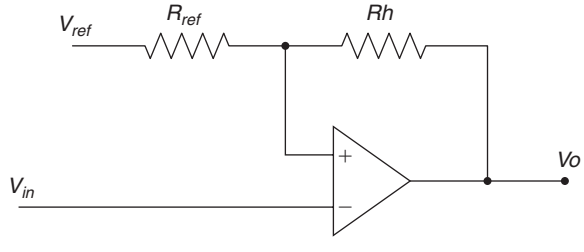


FIGURE 3.15
Positive feedback on an op-amp.

Here is the result due to V_{ref} using the voltage divider rule:

$$V+ \text{ due to } V_{ref} = \frac{V_{ref} * Rh}{Rh + R_{ref}} \qquad \text{Eq. 3.19}$$

Here is the result due to V_o using the voltage divider rule:

$$V+ \text{ due to } Vo = \frac{Vo * R_{ref}}{R_{ref} + Rh} \qquad \text{Eq. 3.20}$$

The result due to both is thus:

$$V+ = (V+ \text{ due to } V_{ref}) + (V+ \text{ due to } Vo) \text{ or,}$$

$$V+ = \frac{V_{ref} * Rh}{Rh + R_{ref}} + \frac{Vo * R_{ref}}{Rh + R_{ref}} \qquad \text{Eq. 3.21}$$

Now insert all the current values and we have:

$$V+ = \frac{2.5 * 100\,\text{K}}{110\,\text{K}} + \frac{4 * 10\,\text{K}}{110\,\text{K}} = 2.64\,\text{V} \qquad \text{Eq. 3.22}$$

Is this circuit stable now? Yes, it is. We have 0 V at $V-$ and 2.64 V at $V+$. This results in a positive error, which, when amplified by the open-loop gain of the op-amp, causes the output to go to the positive rail. This is 4 V, which is the state that we just analyzed.

Now let's change something and see what happens. Let's start slowly ramping up the voltage at $V-$. At what point will the op-amp output change? Right after the voltage at $V-$ exceeds the voltage at $V+$. This results in a negative error, causing the output to swing to the negative rail. And what happens to $V+$? It changes back to 2.275 V, as we calculated above. So how do we get the output to go positive again? We adjust the input to less than 2.275 V. The positive feedback reinforces the change in the output, making it necessary to move the input farther in the opposite direction to affect another change in the output.

The effect that I have just described is called *hysteresis*. It is an effect very commonly created using a positive feedback loop with an op-amp. What is hysteresis good for? you ask. Well, heating your house, for one thing. It is hysteresis that keeps your furnace from clicking on and off every few seconds. Your oven and refrigerator use this principle as well. In fact, the disk drive on the computer I used to write this paragraph uses hysteresis to store information.

One important item to note: The size of the hysteresis window depends on the ratio of the two resistors R_{ref} and Rh. In most typical applications, Rh is much larger than R_{ref}. If the signal at Vi is smaller than the window, it is possible to create a circuit that latches high or low and never changes. This is usually not desired and can be avoided by performing the preceding analysis and comparing the calculated limits to the input signal range.

Now that we have covered the three basic configurations of an op-amp, let's put together a simple circuit that uses them. Here we have a voltage follower, hooked to a comparator using hysteresis, with an LED as an indicator (Figure 3.16). You should build this in your lab to gain an intuitive understanding of what has been discussed. Experiment with feedback changes in all parts of the circuit. Note that you can change the input potentiometers from 5 to 100 K without affecting the voltage at which the comparator switches.

All About Op-Amps

There you have it—the basics of op-amp circuits. With this information, you can analyze most op-amp circuits you come across and build some really neat ones yourself. What about filters, you say! Well, a filter is nothing more than an amplifier that changes gain, depending on the frequency. Simply replace the

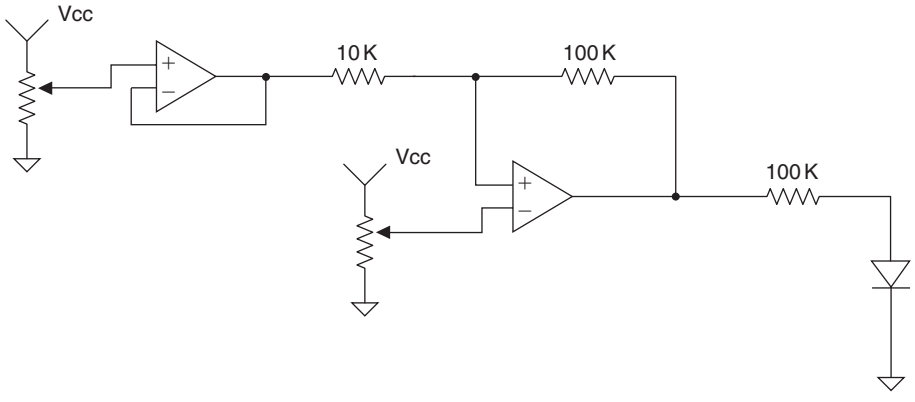


FIGURE 3.16
Simple op-amp circuit for your bench to help you understand both positive and negative feedback.

resistors with a cap or inductor and thus add a frequency component to the circuit. What about oscillators, you say? These are feedback circuits where timing of the signals is important.[17] They still follow the preceding rules. I must reiterate my belief that grasping the basics of any discipline is the most important thing you can do. If you understand the basics, you can always build on that foundation to obtain higher knowledge, but if you do not "get the basics," you will flounder in your chosen field.

Thumb Rules

- Op-amp inputs are high impedance (that means no current flows into the inputs); this can't be said too often, so forgive me for repeating it.
- Op-amp outputs are low impedance.
- V+ = V− only if negative feedback is present; they don't have to be equal if feedback is positive.
- Positive feedback creates hysteresis when properly set up.
- Positive feedback can make an output latch to a state and stay there.
- Positive feedback with a delay can cause an oscillation.
- Op-amps were designed to make it easy, so don't make it hard!

[17] Just the right amount of delay in the feedback and you can get a signal to chase itself back and forth and thus oscillate.

IT'S SUPPOSED TO BE LOGICAL

Binary Numbers

Binary numbers are so basic to electrical engineering that I nearly omitted this section on the premise that you would already know about them. However, my own words, "Drill the basics," kept haunting me. So if you already know this stuff forward and backward, you are authorized to skip this section, but if those same words start to haunt you, as I hope they will, you should at least skim through it.

Binary numbers are simply a way to count with only two values, 1 and 0—convenient numbers for reasons we will discuss later. Binary is also known as *base 2*. There are other bases, such as base 8 (octal) and base 16 (hexadecimal), that are often used in this field, but it is primarily for the reason that they represent binary numbers easily. The common base that everyone is used to is decimal,[18] also known as *base 10*. Think of it this way: the base of the counting system is the point at which you move a digit into the left column and start over at 0. For example, in base 10 you count 0, 1, 2, 3 … 7, 8, 9 and then you chalk one up in the left column and start over at 0 for the number 10. In base 8 you only get to 7 before you have to start over: 0, 1, 2 … 5, 6, 7, 10, 11, and so on. Base 16 starts over at 15 in the same way, but to adhere to the rule of one digit in the column before we roll over into the next digit, we use letters to represent 10 through 15. Table 3.1 shows an easy way to see this relationship.

Note again how the numbers start over at the corresponding base. You might also notice that I started at 0 in the counting process.[19] It should be stressed that 0 is an important part of any counting system, a fact that I think tends to get overlooked. If you think about it, when 0 is included, the point at which base 10 rolls over is the 10th digit and the point at which base 8 rolls over is the 8th digit. The same relationship exists for any base number you use.

So, let's get back to binary or base 2. The first time I saw binary numbers I thought, "Wow, what a tantalizing[20] numeric system; just as soon as you make one move to get where you are going, it is time to start over again." The numbers

[18] You can chalk that up to the fact that we have 10 fingers on our hands. In fact, the ancient Mayans used a base-20 system of counting, presumably due to the fact that they ran around without any shoes.

[19] Here is your chance to giggle at the fact that this new version of my book has a Chapter 0—that is, if you are inclined to think that my dry engineering sense of humor is in fact funny.

[20] Again, it is an odd sort of person who will find a numeric system "tantalizing," but I never said I wasn't odd!

Table 3.1 Decimal and Hexadecimal Numbers

Decimal, Base 10	Hexadecimal, Base 16
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	A
11	B
12	C
13	D
14	E
15	F
16	10
17	11
And so on ...	

go like this: 0, 1, 10, 11, 100 … Again I think a table is in order—see Table 3.2 on next page.

Notice how base 8 and base 16 roll over right at the same point that the binary numbers get an extra digit. That is why they are convenient to use in representing binary numbers. You might also have noticed that decimal numbers don't line up as nicely.

Another pattern you should see in this table is that you hit 20 in base 8 at the same point at which you see 10 in base 16. This makes sense because one base is exactly double the other. Can you extrapolate what base 4 might do?

This leads to another trick with binary numbers. Each significant digit doubles the value of the previous one (just as every digit you add in decimal is worth 10 times the previous one). Let's look at yet another table—see Table 3.3 on the next page.

Table 3.2 Decimal, Binary, Octal, and Hexadecimal Number Comparison

Decimal Base 10	Binary Base 2	Octal Base 8	Hexadecimal Base 16
0	0	0	0
1	1	1	1
2	10	2	2
3	11	3	3
4	100	4	4
5	101	5	5
6	110	6	6
7	111	7	7
8	1000	10	8
9	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F
16	10000	20	10
17	10001	21	11
18	10010	22	12
And so on …			

Table 3.3 Doubling Digits

Decimal	128	64	32	16	8	4	2	1
Binary	10000000	1000000	100000	10000	1000	100	10	1

You can add up the values of each digit where you have a 1 in binary to get the decimal equivalent. For example, take the binary number 101. There is a 1 in the 1s column and in the 4s column. Add 1 plus 4 and you get 5, which is 101 in binary. You might also notice that the numbers you can represent double for every digit you add to the number. For example, four digits let you count to 15, and eight digits will get you to 255. (This causes some of us more extroverted engineers to attempt to become the life of the party by showing their friends that they can count to 1023 with the fingers on their hands. These attempts usually fail.)

All the math tricks you learned with decimal numbers apply to binary as well, as long as you consider the base you are working in.

For example, when you multiply by 10 in decimal, you simply put a 0 on the end, right? The same idea applies to binary, but the base is 2, so to multiply by 2, you simply stick a 0 on the end, shifting everything else to the left. When dividing by 10 in decimal you simply lop off the last digit and keep whatever was there as a remainder. Dividing by 2 in binary works the same way, shifting everything to the right, but the remainder is always 0 or 1—a fact that is convenient for math routines, as we will learn later.

For whatever reason, most electronic components like to manage binary numbers in groups of four digits. This makes hexadecimal (or *hex*) numbers a type of shorthand for referring to binary numbers. It is a good shorthand to know.

In the electronics world, each binary digit is commonly referred to as a *bit*. A group of eight bits is called a *byte* and four bits is called a *nibble*. So if you "byte" off more than you can chew, maybe you should try a "nibble" next time.

Back to the point: Since a hex number nicely represents a nibble, and there are two nibbles in a byte, you will often see two hex numbers used to describe a byte of binary information. For example, 0101 1111 can be described as 5 F or 1110 0001 as E 1. In fact you can easily determine this by looking up the hex equivalent to any nibble using Table 3.2.

To sum things up, binary numbers are a way to count using only two symbols; they are commonly referred to using hex numbers as a type of shorthand notation. When logic circuits came along, the fact that they represented information with only two symbols—on or off, high or low—made them dovetail nicely with binary numbers and binary math.

Logic

One of the most incredible growth industries over the last 50 years has come from the application of electronics to manipulate data based on the principles of *Boolean logic*. Originally developed by George Boole in the mid-1800s, Boolean logic is based on a very simple concept yet allows creation of some very complex stuff.

Let the value 1 mean true, and let the value 0 mean false. In an actual circuit, 1 might typically be any signal between 3 to 5 V, and 0 any signal between 0 to 2.9 V, but what is important in the world of logic is that there are only two states, 1 or 0. The world is black or white. That said, it is no wonder that engineers have so quickly grasped the digital domain. I haven't met an engineer who

doesn't like his world to follow nice, predictable rules. "Keep it simple" is a common mantra, and resolving the world into two states sure does simplify things. It is important to note that at some point in the circuit a decision needs to be made whether the current value represents a 1 or a 0.

During our study of logic we will refer to a description of logic inputs and outputs known as *truth tables*. In these tables, the inputs are generally shown on the left and the outputs are on the right. Some basic components that manipulate logic are called *gates*. Let's start with these basics.

The NOT Gate

This is as simple as it gets. The NOT gate inverts whatever signal you put into it; put in a 1, get a 0 out, and vice versa. Let's take a transistor and make a NOT gate, as shown in Figure 3.17.

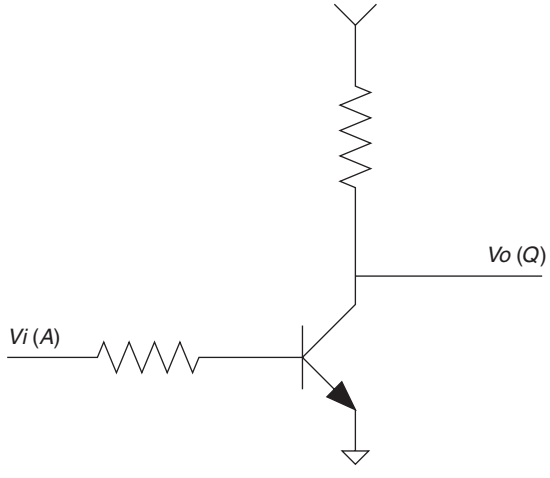


FIGURE 3.17
Transistor NOT gate.

If you put 0V into this, you will get 5V out. If you put 5V into this, you will get nearly[21] 0V out. You have effectively inverted the logic symbol. The NOT gate, also called the *inverter*, is commonly represented by the symbol shown in Figure 3.18. Table 3.4 shows the truth table.[22]

[21]Please note that I said *nearly* 0 volts. The output of this circuit does not quite get all the way to 0, but that doesn't matter as long as the value is below the maximum level for a 0. That right there is the reason digital is so pervasive.

[22]A truth table is a "map" of inputs vs. outputs on a logic device. Kinda makes me wonder what a "lie" table might look like …

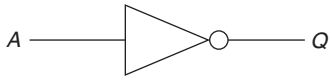


FIGURE 3.18
Inverter or NOT symbol.

Table 3.4 NOT Gate Truth Table

Input A	Output Q
1	0
0	1

The AND Gate

The AND function is described by the rule that all inputs need to be true or 1 in order for the output to be true. If this is true and that is true, this AND that must be true. However, if either is false, the output must be false. It is defined by the truth table shown in Table 3.5.

Table 3.5 AND Gate Truth Table

Input A	Input B	Output Q
0	0	0
0	1	0
1	0	0
1	1	1

We can build this circuit with only a couple of diodes. One way to think of it is that if either input is false, the output will be false—see Figures 3.19 and 3.20. This function is commonly referred to by the symbol.

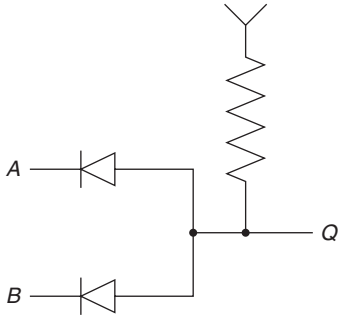


FIGURE 3.19
Diode AND gate.

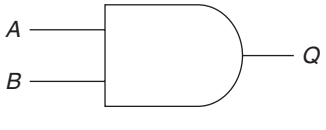


FIGURE 3.20
Another diode AND gate.

The OR Gate

Did you notice that three of the input conditions on the AND gate resulted in a false, or 0, at the output? The OR gate is sort of the opposite, but not exactly. Three of the input conditions result in a true at the output, whereas only one condition creates a 0. If *this* is true OR *that* is true, it only takes one true input to create a true output. Table 3.6 shows the truth table.

Table 3.6 OR Gate Truth Table

Input A	Input B	Output Q
0	0	0
0	1	1
1	0	1
1	1	1

We can make this circuit with diodes too; we just flip them arounds, as in Figure 3.21. The more common OR symbol looks like the one shown in Figure 3.22.

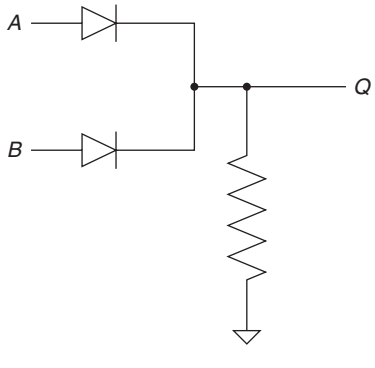


FIGURE 3.21
Diode OR gate.

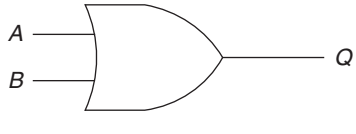


FIGURE 3.22
Most common OR symbol.

That's it—those are the basic gates. There are only three of them. "Now wait a minute," you may be saying, there were a lot more when I had logic circuits in class, weren't there? There are more gates, but they are all built from these three basic gates. If you understand these, you can derive the rest. With that in mind, see if you can make these other logic gates using only the previous three components.

The NAND Gate

NAND means NOT AND, and it is what it says. Invert the output of an AND gate with the NOT gate and you have a NAND gate. Table 3.7 shows the truth table.

Table 3.7 NAND Gate Truth Table

Input A	Input B	Output Q
0	0	1
0	1	1
1	0	1
1	1	0

Let's build one with the basic symbols we have already learned, as shown in Figure 3.23. This gate is so commonly used that it has its own symbol. Note the little bubble on the output, which is used to indicate an inverted signal.

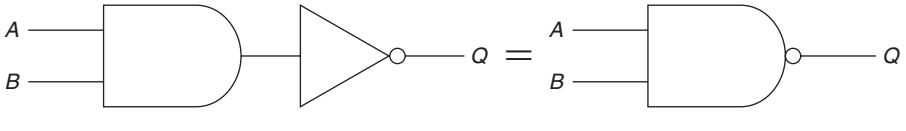


FIGURE 3.23
How to build a NAND gate.

Can you make this gate with basic semiconductors as well? The answer is yes. In fact, you only need two transistors—see Figure 3.24 on the next page.

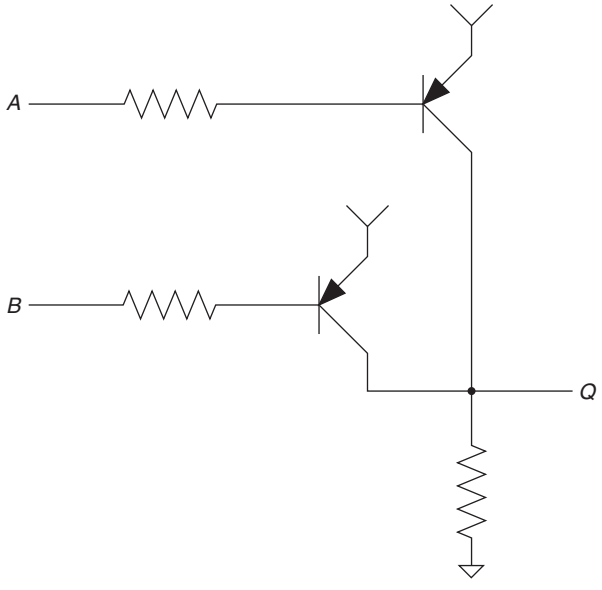


FIGURE 3.24
Simple transistor NAND gate.

The NOR Gate

Yep, you guessed it, this is the NOT OR gate. It is made by inverting the output of the OR gate, just like the NAND gate. Table 3.8 shows the truth table. The NOR gate is an inverted OR gate with a symbol like the one shown in Figure 3.25. Better yet, as Figure 3.26 shows, you can make this gate with only two transistors as well.

Table 3.8 NOR Gate Truth Table

Input A	Input B	Output Q
0	0	1
0	1	0
1	0	0
1	1	0

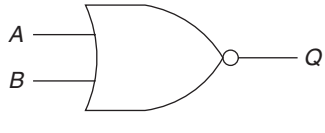


FIGURE 3.25
NOR gate symbol.

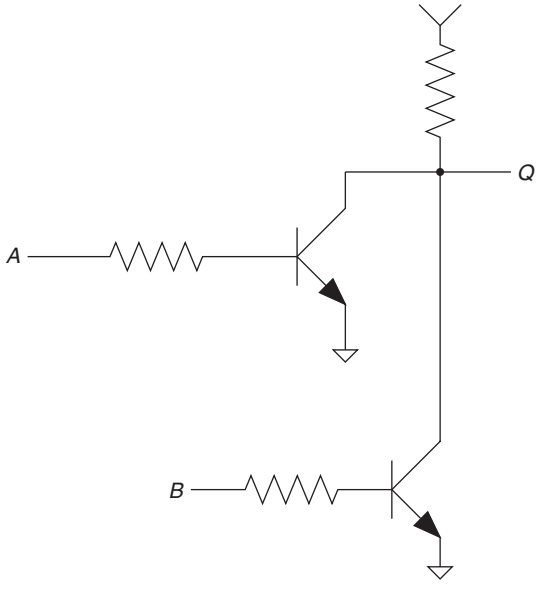


FIGURE 3.26
Transistor NOR gate.

The XOR Gate

XOR means *exclusive or*—see Figure 3.27. In other words, think of it like this: it's true if *this* or *that* is true, but not if both are true. Table 3.9 shows the truth table.

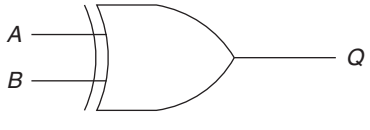


FIGURE 3.27
XOR (exclusive OR) gate.

Table 3.9 XOR Gate Truth Table

Input A	Input B	Output Q
0	0	0
0	1	1
1	0	1
1	1	0

Let's see whether we can make this with basic semiconductor components the some as we did with the other logic circuits, as shown in Figure 3.28 on the next page.

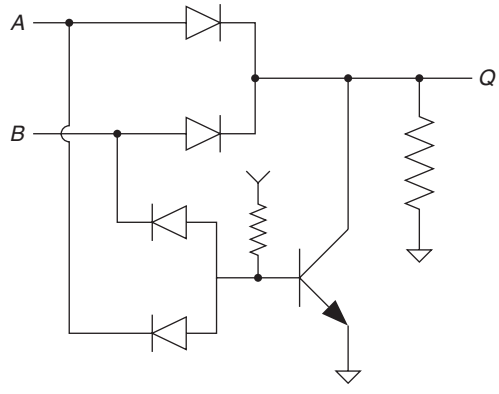


FIGURE 3.28
Diode- and transistor-based XOR gate.

The XNOR gate looks like the one in Figure 3.29. If I have done a good job with my explanations, the function of this gate should be obvious. It is an XOR with an inverted output. Table 3.10 shows its truth table.

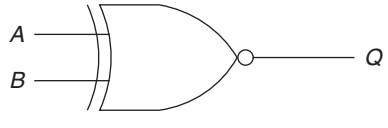


FIGURE 3.29
The XNOR gate.

Table 3.10 XNOR Gate Truth Table

Input A	Input B	Output Q
0	0	1
0	1	0
1	0	0
1	1	1

Adders

As you already know, it is possible to count with these ubiquitous 1s and 0s. The logical extension of counting is math! Joining several of these gates together, we can create a binary adder; string a bunch of these adders together

to add any number of binary digits and, since any number can be represented by a string of those pesky 1s and 0s, we now have the basis of computation. Are you beginning to see how that calculator[23] on your desk works?

Memory Cells

It is possible to use these devices to create what is called a *memory cell*. Figure 3.30 presents a diagram of one.

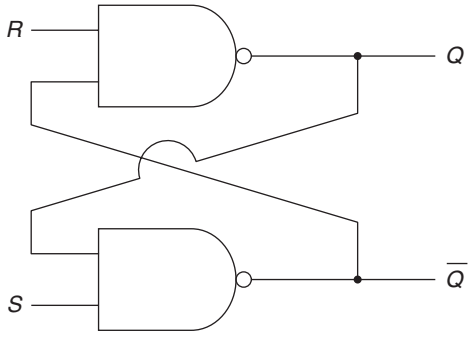


FIGURE 3.30
NAND-based memory cell.

The basic premise is that the cell will retain the state you set it to. Some memory will lose the data that was stored if power is lost; this is known as *volatile memory*. This is like the RAM in your computer. Another category of memory is known as *nonvolatile memory*. In this type the data is retained even when power is removed. An example of this is Flash memory, commonly found in the now-ubiquitous thumb drive.

Now that you have the ability to make a decision, compute mathematical functions, and remember the results so you can make more decisions later, you have the basics of a *Turing machine*. Alan Turing was a cryptographer who laid much of the foundation for computational theory. He described the Turing machine, a system that has an infinite amount of memory, the ability to go back and forth along that memory, and the capability to follow the instructions at any location. Aside from infinite memory, today's computers are as close as anything comes to a Turing machine.

[23] Technically, most calculators use a CORDIC algorithm. It is a slick way to handle things like sine, cosine, and other stuff and still keep the electronics simple. At the end of the day, though, deep down inside that desktop appliance there are still logic elements doing all the work.

From the simple gates that started it all to supercomputers, ever more complex systems are based on these simple logic components. It is no wonder that every new mega-cool processor has a gazillion transistors in it. There is a sort of "in-between" device that is worth mentioning, though, since it will help you grasp the complexities such a simple device can create. It is known as a *state machine*.

State Machines

State machines lie in the realm between discrete logic and microcontrollers. They usually have a clock of some type, memory, and most of the basic parts a micro has; however, they don't need all these parts to operate.

As the name implies, the output of a state machine is a function of the "state" of the inputs at any given moment in time. Often a clock signal of some type is used to determine the moment that these inputs should be evaluated. Memory cells, also called *flip-flops*, are used to store information. A flip-flop reflects the state of the input at the time a clock signal was present. Thus conditions used for evaluation can be stored in memory.

The inputs of a logic element can be detected at three different points in time on the clock signal, falling edge, rising edge, or level detect. The one that is used depends on the part itself; you will need to check that source of all knowledge, the datasheet.

These terms are self-explanatory: Data is assessed when the clock signal rises, falls, or remains level. This makes the timing of the signals important. This importance of timing will come up again as we explore microcontrollers (which are really just hopped-up state machines with a defined group of instructions, but more on that later).

Due to the falling cost of microcontrollers, I believe that purely implemented state machines are going out of fashion these days. When they do appear, they are usually in a *programmable logic device*, also called a *PLD*. Gone are the days of soldering a slew of D flip-flops onto a board and wire-wrapping a circuit together.[24] Even PLDs now have an MCU core that you can cram in there for general computing needs.

In conclusion, Boolean logic is the foundation of all things digital. It is a relatively simple concept that can do some very complex things. Ours is clearly

[24]Have you noticed that the older you get, the more natural it seems to enter a state of blissful reminiscing? What could be the evolutionary benefit of that?

becoming a digital world. When was the last time you saw the latest widget marketed as the coolest new "analog" technology?

Thumb Rules

- Every significant digit you add in binary doubles the value of the previous digit.
- A bit is a single piece of information with only two states, 1 or 0.
- There are 4 bits to a nibble and 8 bits to a byte.
- 1 is true, 0 is false.
- Always look at the truth table.
- At some point in the circuit, a signal is considered either high, 1, or low, 0; what it is depends on the thresholds of the part.
- Timing is very important in setting up more complex logic circuits.

MICROPROCESSOR/MICROCONTROLLER BASICS

This is one of the most rapidly changing fields in the electronics industry. You can purchase microcontrollers today with only six pins with just a few lines of memory at a cost of 25 cents and for just a few bucks more, high-end embedded processors that just a few years ago would have been labeled supercomputers. All this from the few semiconductor types we have discussed. I will not try to cover specific processors since there are libraries of books dedicated to understanding particular micros. Instead I will try to cover some fundamental rules that can be applied in general.

Add a bunch of logic gates together and mix with some adders, instruction decoders, and memory cells. Hook it all up to some input/output pins, apply a clock source, and you get a microcontroller or microprocessor.

These two devices are very similar, and you will hear the names used somewhat interchangeably. Generally, however, the microcontroller is more all inclusive, with all the elements it needs to operate included in one piece of silicon, typically making them a little (but not much) more specialized. The microprocessor by contrast needs external memory and interface devices to operate. This makes it more open ended, allowing memory upgrades without changing the chip, for example. As this area of technology has progressed, the line of distinction

between these two components has blurred considerably. Hence, much of the design philosophy needed to make the most of these devices is the same.

What's Inside a Micro?

It might seem like magic, but all that is inside a microcontroller is a whole lot of transistors. The transistors form gates, and the gates form logic machines. Let's go over some of the parts that are in a micro.

INSTRUCTION MEMORY

I would call instruction memory ROM, or read-only memory, but these days there are a lot of micros that can write to their own instruction memory. This can be programmable memory, hard coded, Flash, or even an external chip that the core reads to get its instructions. The instructions are stored as digital bits, 1s and 0s, that form bytes that represent instructions.

DATABUS

The databus is the backbone of the micro, the internal connections that allow different parts of the micro to connect internally. Virtually everything that happens inside a micro will at some point move through the databus.

INSTRUCTION DECODER

An instruction decoder is one of those logic type circuits. It interprets the instruction that is presented and sets the corresponding tasks into motion.

REGISTERS

Registers are places to store data; they are literally the memory cells that we discussed earlier. This is the RAM inside the micro. It is the scratch pad for manipulating data. It can also be accessed on an external chip in some cases.

ACCUMULATOR

An accumulator is a type of special register that usually connects directly to the arithmetic logic unit (ALU). When a math function is performed on a piece of data in the accumulator, the answer is left in the accumulator; hence it *accumulates* the data. On a lot of the newer micros, nearly any register can be used in a similar manner.

ALU

The arithmetic logic unit, or ALU, is a part that can perform various mathematical and logic operations on a piece of data.

PROGRAM COUNTER

The program counter keeps track of where the micro is in its program. If each piece of memory were a sheet of paper with a number on it, the program counter is the part that keeps track of the number on the sheets. It indexes or addresses which sheet it is on.

TIMER COUNTERS

Timer counters are useful for creating a structure for your code to operate in. Sometimes called *real-time clock counters* (RTCC), they are counters that usually can run from an independent source. They will "tick" at whatever interval you set them up to tick, without any other intervention. Sometimes they can be hooked up to external clock sources and inputs. Usually they can be set to generate an interrupt at a preset time.

INTERRUPT

Not exactly a specific hardware component in a micro, the interrupt is so important that it warrants mention. An interrupt is a monitoring circuit that, if triggered, makes the micro stop what it is doing and execute a piece of code associated with the interrupt. These signals can be generated by internal conditions or external inputs. Typically only certain pins can drive interrupts.

MNEMONICS AND ASSEMBLERS

We humans, unlike machines, have a tough time remembering endless streams of binary data. Even trying to remember all the hex codes for a micro is very difficult. For this reason *mnemonics* were invented. Mnemonics are nothing more than code words for the actual binary data stored in the instruction memory.

An assembler takes these code words and changes them to the actual data, creating a file that is then copied into the instruction memory. This differs somewhat from compilers used to compile code that you write for a computer. The compiler takes a code language such as C, for example, and creates code that will run on the computer. However, the compiler will handle tasks such as addressing memory without any need for you to worry about it, unlike an assembler. This is why they are called higher-level languages. Assembly language, as it is called, works directly with the hardware that the chip is hooked up to.

There are a lot of micros these days that have C assemblers, allowing you to use a language you are familiar with to write code for your micro. However, use caution with this approach. It is possible to lose a lot of efficiency this way. I know of one case where a micro with 4 K of memory was being used to control

an electric toothbrush. The developers coding in C kept coming back for micros with more memory because they couldn't get their code to fit. Once it was written in assembly, the whole thing took about 500 bytes of code. This is an extreme case. I'm sure there are much more efficient designs out there using C. Just be sure you have an idea of what your code is turning into.

Structure

The various ways you can structure your code are as infinite as numbers themselves. There are some basic methodologies that I wish I had been taught before someone handed me a chip and an application note in the lab.

Most microcontrollers only do one thing at a time.[25] Granted, they can do things very fast so as to appear to be multitasking, but the fact is, at each specific instruction only one thing is being accomplished. What this means is that timing structure can have a huge effect on the efficiency of a design.

Consider this simple problem. You have a design where you need to look at an input pin once per second. One way of doing this is as follows (note the use of darrencode, a powerful and intuitive coding tool. Too bad it doesn't run on any known micro!):

Initialization
Clear counters
Setup I/O
Sense input
    Read pin
    Store reading
Delay loop
    Do nothing for 1 microsecond
    Jump to Delay loop 100,000 times
Delay done
    Jump to Sense input

There is a slight problem with this method that you might have already noticed. The processor spent the whole time waiting for the next input, doing nothing. This is fine if you don't need the chip to do anything else. However, if you want to get the most out of your micro, you need to find a way to make it

[25] Due to Moore's Law, this is becoming a less true statement these days. Today readily available multicore processors are out there that can do more than one thing at a time. The same general rules apply; you just have some additional ability to consider.

do something else while you wait and come back to the input at the right time. The best way to do this is with *timing interrupts*.

An interrupt is just what it says. Imagine you have an assistant that you have told to watch the clock and remind you right before 5:00 p.m. that you need to go to that important meeting. You are hard at work when your assistant walks in and *interrupts* you to let you know it is time to go. Now if you are as punctual as one of these chips, you drop whatever you are doing and go take care of business, coming back to your task at hand after you have taken care of the interruption. In micro terms this is known as servicing the interrupt.

Most micros have a timer that runs off the main clock which can be set to trigger an interrupt every so often. Let's solve the previous problem using interrupt timing and see how it looks:

```
Initialization
Setup Timer Interrupt to trigger every 1 microsecond
Clear counters
Setup I/O
Main loop
    Calculate really fast stuff
Tenth second loop
    Check tenth second flag
    Jump to End tenth if not set
    Do more tasks
    Call some routines
    End tenth
Second loop
    Check second flag
    Jump to End second if not set
    Read pin
    Store reading
End second
    Jump to Main loop
Timer Interrupt
    Increment microsecond counter
    If microsecond count equals 10,000
        set tenth second flag
```

```
        increment tenth counter
        clear microsecond count
    Else clear microsecond flag
If tenth count equals 10
    set second flag
    clear tenth count
end interrupt
```

One thing to note is that you don't want to put a lot of stuff to do inside the interrupt. If you put too much in there you can have a problem known as *overflow*, where you are getting interrupted so much that you never get anything done. (I'm sure you have had a boss or two who helped you understand exactly how that feels.) In the darrencode example, the only thing that happens in the interrupt is incrementing counters and setting flags. Everything that needs to happen on a timed base is done in the main loop whenever the corresponding flag is set.

The cool thing is that now we have a structure that can read the input when you need it to and still have time to do other things, such as figure out what that input means and what needs to be done about it. This structure is a rudimentary operating system. In my case, I like to call it darrenOS. Feel free to insert your name in front of a capital O and S for the timed code you create on your next micro. (*Insert your name here*)OS is a free domain, and I promise you won't get any spyware using it!

The biggest downside to this type of structure, in my opinion, is the added complexity in understanding how it works. The first example is straightforward, but as you step through the second example, you might notice it is a bit harder to follow. This can lead to bugs in your code simply because of the increased difficulty in following the logic of your design. There is nothing wrong with the first example if you don't need your micro to be doing anything else. However, the timing structure in the second design is ultimately much more flexible and powerful. The trade-off here is simplicity as well as limited code execution for complexity and the ability to get more out of your micro.

Some of you out there with some coding experience might now be saying, "Why not just run the input pin you need to check into an interrupt directly and look at it only when it changes?" That is a good question. There are times when this interrupt-driven I/O approach is clearly warranted, such as when extreme speed in response to this input is needed. However, in any given micro, you have only a few interrupts available. If you did that on every I/O

pin, you would soon run out of interrupts. Another benefit of this structure is that it will tend to ignore noise or signal bounce that sometimes happens on input pins that are connected to the outside world.

Some Slick Math Routines

It's not too hard to write a routine to multiply or divide. It can be difficult, however, to write *good* multiply and divide routines. Some of the characteristics of good routines are that they are short and concise and that they consistently use as little memory as possible.

I've talked with students and other professionals and asked them how they would write multiply and divide routines. Remember, you only get to use add, subtract, and other basic programming commands in these small micros that are so cost-effective. The most common approach that engineers come up with is the same method that I first came up with when I tackled the problem. The following is an example.

> We want to multiply two numbers A * B:
>
> 1. Result = 0
> 2. If (B = 0) Then Exit
> 3. Result = Result + A
> 4. B = B − 1
> 5. If (B = 0) Then Exit Else GOTO 3
>
> We want to divide two numbers A/B:
>
> 1. Result = 0
> 2. Remainder = A
> 3. If (B < A) Then Exit
> 4. Remainder = Remainder − B
> 5. Result = Result + 1
> 6. GOTO 3

These routines will work and they have some advantages: You use very little RAM or code space, and they are very straightforward and easy to follow. However, they have one significant disadvantage: These routines could take a long time to execute.

The multiplication routine, for example, would execute quickly if B = 3, but if B = 5,000, the routine would take much, much longer. The divide routine runs into the same problem because the ratio of A to B becomes very large. Anyone

who spends their days trying to squeeze performance out of the bits and bytes world knows that this is a no-no. Routines like this would cause you to spend all your time trying to find out why the chip resets, because of watchdog timers expiring when a big number gets processed.

Fortunately, there is a better way. I was shown the following methods and I pass them on to you as useful tools. It isn't a great secret; you just need to get out of that old mundane base-10 world and think like a computer.

The binary world has one reoccurring advantage: When you shift numbers to the left once, you multiply that number by 2. If you shift numbers right once, you divide by 2. Not too hard, right? After all, we've followed a similar rule since we were little in our decimal world. Shift one digit to the left and we multiply by 10, shift 1 digit to the right and we divide by 10.

Using this simple rule with addition and subtraction, we can write multiply and divide routines that are accurate, expandable, use very little code or RAM, and take approximately the same number of cycles no matter what the numbers are. The examples that follow will be byte-sized for simplicity, but the same pattern can be used on operands of any size. You just need the register space available to expand on this idea.

Multiplication

Let's start with two numbers A * B. For this example, we will say that A = 11 and B = 5.

In binary, A = 00001011 and B = 0000101

When multiplying two-byte sized numbers, you should know that the result can always be expressed in two bytes. Therefore, RESULT is word sized, and TEMP is word sized. COUNT needs only to be one byte.

1.	RESULT = 0;	This is where the answer will end up
2.	TEMP = A;	necessary to have a word-sized equivalent for shifting
3.	COUNT = 8;	This is because we are multiplying by an 8-bit number
4.	Shift B right through carry;	Find out if the lowest bit is 1.
5.	If (carry = 1) then RESULT =	RESULT + TEMP
6.	TEMP = TEMP + TEMP;	Multiply TEMP * 2 to set up for next loop
7.	COUNT = COUNT − 1	
8.	If (COUNTER = 0) then exit else GOTO 4	

Look at the mechanics of this. As we rotate or shift B through carry each time, we are simply moving left in B each time through the loop and deciding whether B has a 1 or a 0 in that location. (Remember, moving left is multiplying by two.) At the same time, we are shifting TEMP left each time since the binary digit we are checking in B is double the magnitude it was the previous time through the loop.

Then all that is left to do is add the TEMP value if the value of the binary digit in B is 1, or don't add it if B has a 0 in that location. By the time COUNT = 0, you have the final result in RESULT. The loop works the same way no matter how large your numbers are. The subroutine has a somewhat small range of possible machine cycles that it takes and still remains compact and uses a minimal amount of RAM.

Let's look at our example problem in a table form, as shown in Table 3.11; by the time it reaches step 8 the operation is complete. (Note that *x* = Don't care.)

Table 3.11 Example Problem

Loop Count	RESULT	B	TEMP	COUNT
1	00000000 00001011	x0000010	00000000 00010110	7
2	00000000 00001011	xx000001	00000000 00101100	6
3	00000000 00110111	xxx00000	00000000 01011000	5
4	00000000 00110111	xxxx0000	00000000 10110000	4
5	00000000 00110111	xxxxx000	00000001 01100000	3
6	00000000 00110111	xxxxxx00	00000010 11000000	2
7	00000000 00110111	xxxxxxx0	00000101 10000000	1
8	00000000 00110111	xxxxxxxx	00001011 00000000	0

Division

Now that multiplication is clear, division is simply multiplication in reverse. Let's take the numbers A = 102 and B = 20 and perform A/B. In binary: A = 01100110 B = 00010100.

Since we are dealing with integers, we know that A/B has a RESULT less than or equal to A. Therefore, RESULT is one byte, and REMAINDER is one byte. TEMP is two bytes.

1. RESULT = 0 ; This is where the answer will end up
2. REMAINDER = 0; This is for the remainder
3. COUNT = 8 ; This is because we are dividing by an 8-bit number
4. RESULT = RESULT + RESULT
5. Shift A left through carry
6. Shift REMAINDER left through carry
7. If REMAINDER >=B then RESULT = RESULT + 1 and REMAINDER = REMAINDER − B
8. COUNT = COUNT − 1
9. If (COUNTER = 0) then exit else GOTO 4

This might seem somewhat foreign, but it's really the same type of division that you've always known. First we look at how many digits in the top part of A we need before B will divide into those digits. Once we have the number of digits, we subtract that division and then continue. Follow through the table with our example numbers and see if it becomes clear.

Let's look at our example problem again as in Table 3.12; just like before, by the time it reaches step 8 the operation is complete.

Table 3.12 Another Example Problem

Loop Count	A	RESULT	REMAINDER	COUNT
1	1100110x	00000000	00000000	7
2	100110xx	00000000	00000001	6
3	00110xxx	00000000	00000011	5
4	0110xxxx	00000000	00000110	4
5	110xxxxx	00000000	00001100	3
6	10xxxxxx	00000001	00000101	2
7	0xxxxxxx	00000010	00001011	1
8	xxxxxxxx	00000101	00000010	0

Slick, Isn't It?

There are always several ways to do things, and I would never say to you that these are the best math routines for all situations. However, they are very flexible and easy to use. They can easily be adapted for 16-bit, 32-bit, 64-bit, or higher math and still work just as well.

The time that it takes for the math to execute depends on the size of the operands in bits, not the actual value of the operands, giving you more or less consistent time for the routine—a very desirable trait.

Get to Know Your I/O

One of the most important pages of the datasheet for any micro is the section that covers the I/O, or the input and output pins. You should be able to answer some simple questions about the I/O of your micro. For example, how much current can the output source? How much can it sink?

Often I have had a problem getting a micro to work as I expected it to, pouring over the code trying to figure out what went wrong, only to find out that I didn't understand the limitations of the I/O pins. Don't ever assume that all I/O is the same.

Knowing what your I/O is and how it works makes you infinitely more valuable as a programming resource. It sets apart the men from the boys[26] in the embedded programming world.

These are some things you should know about input pins:

1. What is the input impedance?
2. Is there an internal pull-up or pull-down resistor?
3. How long does a signal need to be present before it can be read?
4. How do you set it to an input state?

The last might seem like a strange question, but I once worked with a micro that had an input that was an input only when you wrote a high to the output port. If you wrote a low to the output port, it became an output. It was a kind of funky open-drain I/O combination. Here are some things you should know about output pins:

1. What is the output impedance?
2. How much current can it sink?
3. How much current can it source?
4. How long will it take to change state under load?
5. How do you configure it to be an output?

Did you notice the timing questions? Timing, especially when accessing stuff like external memory, is important. You need to know how fast you can get the signal out of the micro and how long it takes the micro to see the signal. With timing problems, your design might work great on a few prototypes only to manifest all sorts of odd behavior later in production on a percentage of the production run. To sum it up, it is very important to understand what your I/O can and can't do.

[26] Or "women from the girls"; in today's world you have to be politically correct even in your euphemisms.

Where to Begin

Many times I have seen an engineer (myself included) work for hours, even days, on his or her code only to program a micro, sit back and … watch it do nothing. You wiggle some wires, check power and … still nothing. Where do you go from here?

Sometimes the best thing you can do is try to get the simplest of operations going—something like toggling an LED on and off every second. If you use the timing structure that we discussed earlier, getting an LED to flash will verify several things:

- You will know that your clock is going.
- You will know that your interrupts are working.
- You will know that your timing structure is in place.

If you do not have an LED to flash, hook up a meter or a scope to an output pin and toggle that signal. Once you have this LED that you can toggle on and off at will, you can begin adding to your code base the more and more complex routines you will need for a particular project. The moral of the story is, Don't try to get all your code functional all at once. Try to do some simple operations (so simple they are probably not even in the functional specification) first. Once you get some simple things down, the more complex stuff will come much easier. It is easier to chase down code-structure problems on a single LED than it is on a 32-bit DRAM data interface!

Thumb Rules

- Understand the main components of the micro.
- There are times when coding in a lower-level language is preferable.
- Creating a timing structure is a way to get more out of your micro.
- Don't be afraid to use darrencode or darrenOS or create your own code and OS to help you better understand what is going on.
- Know your I/O.
- Start by simply toggling an LED with your code and go from there.
- Have a smart brother who thinks in binary.[27]
- Do simple things with your code first. Flash an LED.

[27]The part on math routines is adapted with permission from an article my brother Robert Ashby wrote several years ago. Pretty slick, isn't it! He has a book on Cypress PSoC micros that I highly recommend if you want to use that chip. Next to the guys who designed the part, he knows more than anyone I know about the ins and outs of that dog! The book is *Designer's Guide to the Cypress PSoC* (Elsevier, 2005).

INPUT AND OUTPUT

The whole point of these devices is to put something in just to get something out. So it stands to reason that it's worthwhile to devote a few words to this topic.

Input

Like the robot in the movie *Short Circuit*, all the circuits you will ever design will need input. Let's review some common input devices and a little info about them.

There are a few different ways to get these signals into your MCU. One method is via an interrupt. You can hook a signal into a pin that can interrupt the micro. When it does, the micro decides what to do about it and moves on. This has the advantage of getting immediate attention from the micro.

Another way to monitor an input line is to use a method called *polling*. Polling works the same way those annoying telemarketers[28] do. They decide when to call you and ask for information. In the same way, the micro decides when to look at a pin and polls the pin for information.

A third way, becoming more and more common with even the smallest micro, is to take an analog reading. By nature this is a polling operation. You need to tell the A/D when to take a reading. In some cases, however, a pin can be set up as a comparator, and the output of that comparison can drive an interrupt. With that in mind, let's take a look at some common input devices.

SWITCHES

Probably the most basic input device you will encounter is the *switch*. A switch is a low-impedance device when it is closed and the perfect high-impedance connection when it is open. This is because an open switch is disconnected and a closed switch is about as close as you can get to a perfect short. This is important to note because if you are connecting a switch to a high-impedance port on your MCU, when it is open you will have a high impedance[29] connected to a high impedance. This is a sure way to get some weird results. The higher the impedance, the more easily disrupted the signal. To combat this, use a pull-up or pull-down resistor.

---

[28]This is assuming they are the micro. If you are the micro, I guess they would be an interrupt.

[29]When you see the words *high impedance*, think high resistance to both DC and AC signals.

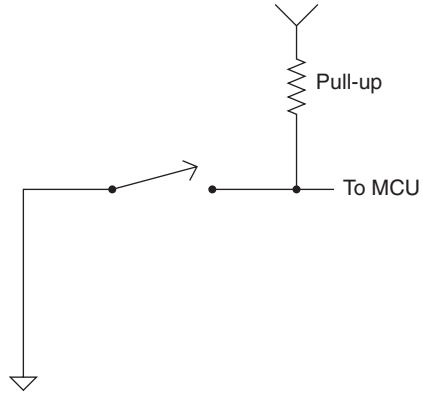


FIGURE 3.31
Switch with pull-up.

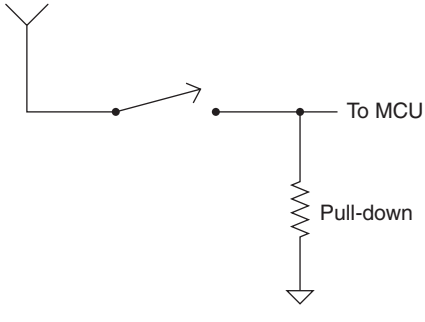


FIGURE 3.32
Switch with pull-down.

A pull-up (or -down) resistor is used to make sure that when nothing else is going on you get a known state on your input line, as shown in Figure 3.31. If you have a switch that when pressed connects the line to ground or reference, use a pull-up resistor to "pull" the signal "up" to *Vcc*. For the opposite situation, use a pull-down resistor,[30] as shown in Figure 3.32.

Generally it is better to poll a switch input than to let it trigger an interrupt. This is due to a phenomenon called *switch bounce*. Being mechanical in nature, a switch internally has two points that come in contact with each other. As they

[30]The value of the resistor in a pull-up or pull-down circuit can be a bit of a tradeoff. The higher the value, the easier the signal will be disrupted by noise; the lower the value, the more current will be used when the switch is closed. You will need to balance those efforts to optimize performance. A good place to start is about 10K.

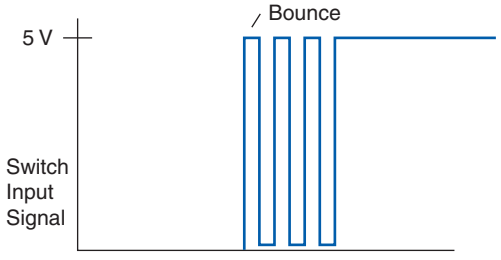


FIGURE 3.33
What happens on a signal line when a switch bounces.

close, it is possible for them to bang open and shut a few times before they close all the way. The contact actually bounces a few times. The input signal to the micro looks like the diagram in Figure 3.33.

If this is an interrupt-driven system, you can see what might happen. Every time the signal goes high, an interrupt is tripped in the micro. When you really only wanted a single action to occur from the switch closing, you might get five or six trips of the interrupt. If you poll this line, you can determine the frequency of the bounce and essentially overlook this problem by checking less often than the frequency of the bounce.[31] Another way to add some robustness is to require two polled signals in a row before you consider the switch closed. This will make it difficult for glitches or noise to be considered a valid input.

TRANSISTORS

Because of the ubiquitous usage of the transistor, it is likely that you will need to interface to it as an input device at some time or another. Like the switch, the transistor is low impedance when it is on and high impedance when it is off, necessitating the need for a pull-up or pull-down resistor. Which one you need depends on the type of transistor you are reading. (See the beginning of this chapter.) Generally you want a pull-up for an NPN type and a pull-down for a PNP type.

PHOTOTRANSISTORS

A cousin to the transistor is the *phototransistor*. This is a transistor that responds to light, often used to detect some type of movement, such as an encoder on the shaft of a motor.

You should treat it the same way as a regular transistor. Note, though, that phototransistors have a gain or beta that can vary much more than a regular

[31] Another way to deal with this is to filter the input with a capacitor.

transistor. You will need to account for that in your design. Another thing you should check with these transistors is their current capability. Usually they won't sink nearly as much current as the basic plain old transistor will, so don't put too much of a load on them.

HALL OR MAGNETIC SENSORS

Hall or *magnetic sensors* are devices that can sense the presence of a magnetic field. They come in all types and flavors, from items called *reed switches* (little pieces of metal in a tube that close when near a magnet) to ICs that can output an analog or digital signal. You will need to look at the output specs on these parts to determine how to set them up. For example, the reed switch you treat like a switch (yes, it can "bounce," too) whereas the hall device might have a transistor output and need a different setup.

DIGITAL ENCODERS

A cousin to the switch, a *digital encoder* switches lines together as you rotate the knob. Like the switch, you will need pull-up or pull-down resistors to ensure reliable readings.

OTHER ICs

There are a multitude of other chips out there from which you can get signals. One thing that is important when talking to other chips is timing. Often you activate the chip you are talking to with an output signal, and then you look at the data coming back. A memory chip is an example of this. You present the address on the address pins and then grab the data from the data pins. One thing you need to consider is the time it takes for the chip to respond to this command. Every digital IC has a response time or propagation delay for it to respond to a signal. You need to make sure you wait long enough for the signal to be present before you try to get it. If there is more than one IC between you and the chip you are talking to, you need to add those delays in as well. Don't just put it together and see if it works without checking this out. It is not uncommon for a chip to be faster than the spec, so one in the lab might work fine, yet when you get into production you will see a seemingly random failure that defies explanation.

INPUT SPECS

Before we move on to analog inputs, there is an important thing to consider when we're dealing with digital inputs. Every micro has input specifications known as *thresholds*. These are the minimum and maximum voltages a signal must reach to be considered a high or a low. You need to make sure that your signal gets above

the maximum and below the minimum. If it spends any time in between, even if it seems to be working right, you can be sure it will cause you trouble down the road. Just remember, between those two values you can't be sure what the micro will consider the signal to be. You won't know if it is a high or low; the micro will resolve it as one or the other. You just can't be sure which one.

POTENTIOMETERS

Potentiometers (also called *pots*) are a type of variable resistor with three connections, commonly called *high*, *wiper*, and *low*. Measuring between pins high and low, you will see a resistor. The wiper is a connection that as it moves touches the aforementioned resistor at various locations. Figure 3.34 shows a symbol of one.

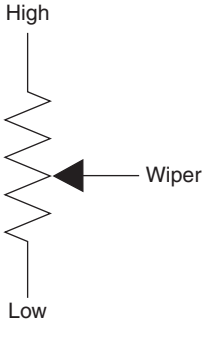


FIGURE 3.34
Diagram of a potentiometer.

If you hook the input voltage to high, wiper to the output, and low to ground, you have nothing more than the voltage divider that we learned about earlier. What is more convenient about the pot is that this voltage divider is easily adjustable by the turn of a knob. If you tie the wiper to one end or the other as shown in Figure 3.35, you will have created a variable resistor that changes as

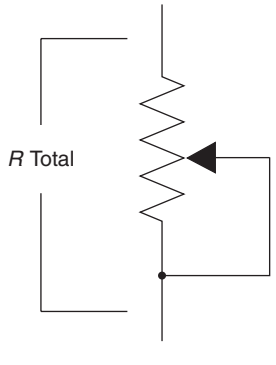


FIGURE 3.35
Potentiometer made into variable resistor.

you move the knob. These are used in myriad ways, to adjust values in a circuit (one without a micro, if you can believe it!) or to tune a device into the correct operation and many other cool things. As it relates to an MCU you might find yourself hooking one of these up as a slick way to dial a value on your project. Commonly you will read these with an A/D input.

Generally, pots have a large tolerance, changing by as much as ±20% in resistance, high to low. However, if used in a voltage divider configuration, this variance is canceled out considerably. This is because while the overall resistance changes, the percentage of resistance for a given position of the wiper doesn't vary nearly as much.

Analog Sensors

Thermal couples, photodiodes, pressure sensors, strain gauges, microphones, and more—a plethora of analog sensors are available. There are so many options that there is no way to cover them all, but here are some good guidelines for using various sensors.

Grounding

Where does the sensor ground go? Dealing with analog sensors requires paying attention to the ground as well as the power source for the sensor. Often the signal line will come right back to the chip reading it, but the ground or power leg might run past multiple ICs before getting to the corresponding pin on the chip reading the signal. This allows currents from all those other ICs to interfere with the current from the sensor. If your sensor is looking at some small signals such as a strain gauge or the like, this can be a bad thing.

Bad: Ground currents from ICs cause noise on the sensor signal—see Figure 3.36.
Good: Traces go back to the chip, keeping the A/D reference where the A/D input is, as shown in Figure 3.37.

Sensor Impedance

What is the output impedance of your sensor? If this is too high with respect to the load[32] it is hooked up to, it might not change the signal in the way you expect. You might need to buffer the sensor so that it is not affected by loading.

[32]This is another place to put to work those estimation skills from Chapter 1. If you have a sensor with a 1-K output impedance, it wouldn't work well to run a 1-K load. Think of it in terms of ratios: Keeping your load higher than 100 K would give you a 100:1 ratio of the output to the load, keeping the amount that could affect it at less than 1%.

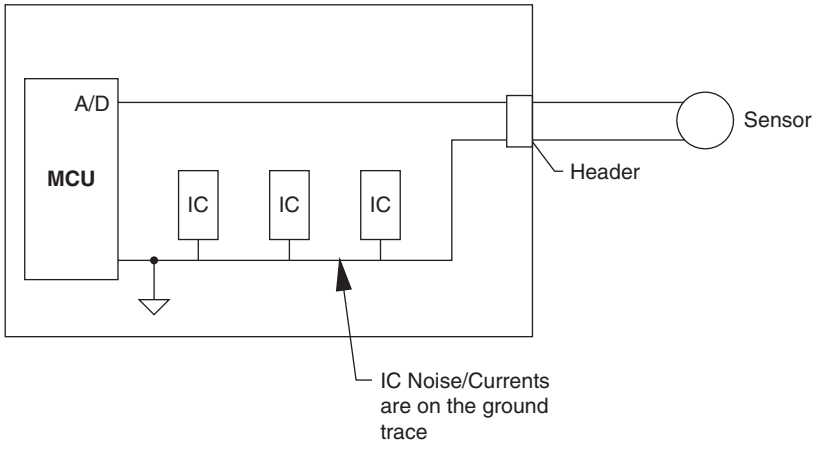


FIGURE 3.36
Poor analog ground layout.

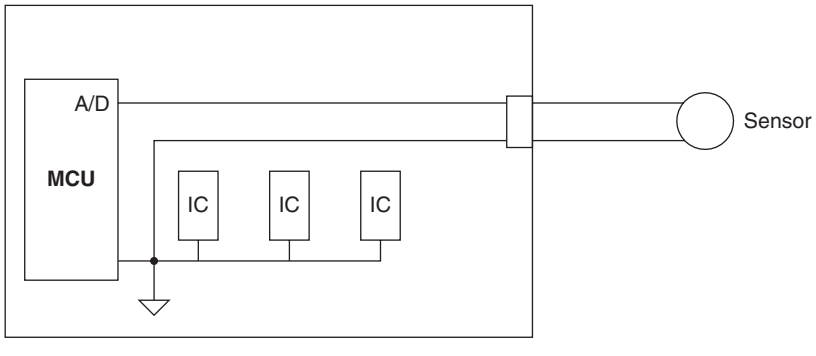


FIGURE 3.37
Much better analog ground layout.

Input Impedance

Most A/D converters have some type of input impedance, usually significantly lower than a digital input. A digital input is often 5 to 10 M ohms of impedance, whereas an A/D may be 100 K ohms. Get to know your input impedance, and make sure it is enough higher than the sensor output impedance so that it's not an issue. A ratio of at least 100:1 is a good place to start. That means that if your A/D is 100 K and your sensor has less than 1 K output impedance, you will have a maximum error of 1%; if that is acceptable in your design you are probably okay.

Output

There are numerous devices that you can output a signal to. We will cover a few of them here. Let's start with some common indicators and displays. Two that are the most common these days are the LED and the LCD.

LEDS

LED stands for *light-emitting diode*. LEDs need current to drive them. Too little and you won't get any light, too much and they will fail, so you typically need a series resistor. How much current is needed depends on the type of LED, but 20 mA is a common normal operating current. LEDs are current-driven devices; this means that their brightness depends on the amount of current flowing through them (not the voltage drop across them). This also means that you can control the brightness by changing the series resistor (Figure 3.38).

An important thing you should consider when driving an LED with a micro is the output capability of the chip. Does the output pin have the ability to source enough current? Can it sink enough current? There are plenty of micros out there that can sink current into a pin but can't source it. For this reason I will typically drive an LED by sinking it—see Figure 3.39.

Do you see how the current flows into the micro? You need to make sure the output pin can handle it! Also, take note that the current flows out of the ground pin on the micro and back to the source.

LEDs have a voltage drop across them, just like the diode that we have already learned about. The new cool blue and white ones are quite a bit higher than

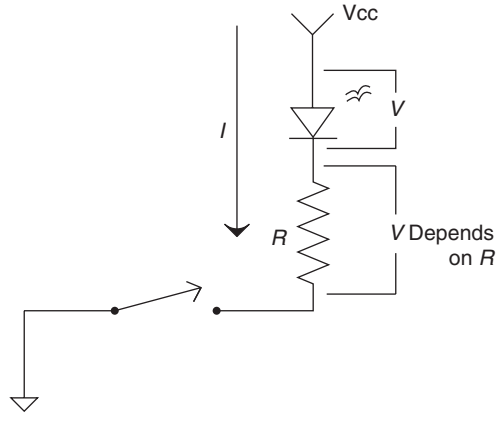


FIGURE 3.38
Switch-controlled LED circuit.

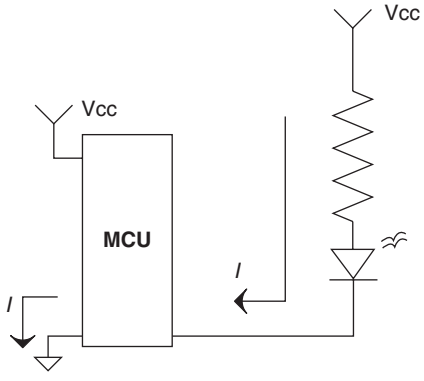


FIGURE 3.39
Diode controlled by MCU.

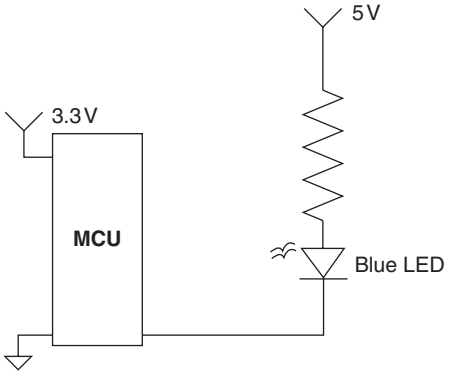


FIGURE 3.40
Less robust way to control a 3.5-V LED with a 3.3-V MCU.

the ones I was raised on. Red, green, and yellow LEDs are around 1.0 to 1.5 V, whereas the blues can easily be 3.5 V.

Figure 3.40 shows a way that you might consider driving one if your MCU has only 3.3 V available as a supply. I wouldn't recommend it, though, because it has a potential problem. Do you see what it is?

The problem with this circuit comes when you try to use the less-expensive, older red/yellow/green diodes. With a smaller voltage drop, current might still flow if the output pin is at a high of 3.3 V and the other end of that diode is at 5 V. Do the math: That would leave 1.7 V across the resistor and the diode, enough to turn it on, albeit weakly in most cases.

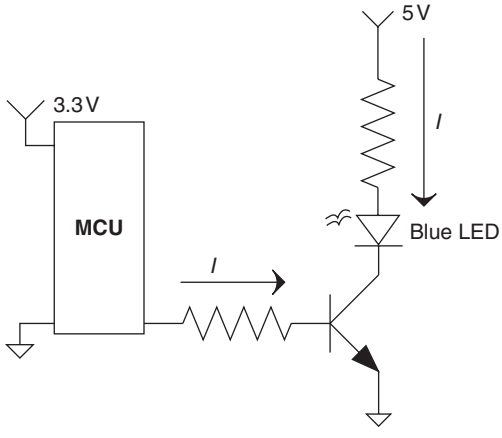


FIGURE 3.41
More robust way to control a 3.5-V LED with a 3.3-V MCU.

Figure 3.41 shows a better way to drive a blue LED under the same constraints.

The moral of the LED story is pay attention to the voltage drop needed to get current moving through it.

Well, enough of the pretty blinky lights; let's examine something that is more fluid.

LCDS

LCD stands for *liquid crystal display*. The liquid crystal in an LCD is a material that responds to an electric field—see Figure 3.42. Applying an electric field to either side of the crystal will make the crystal molecules line up in a certain direction. If you get enough of these crystals lined up, light will be blocked from passing through it.

If you leave an LCD biased for too long, the liquid crystal will permanently twist and you won't be able to twist it back. It is like the crick in your neck that you get from sitting in front of the computer too long. If you don't get up and move a bit every so often, you will tend to stay that way. Though that's good entertainment for fellow employees, a little motion will save you the pain.

The same philosophy works with LCDs. Every so often, reverse the polarity on the LCD and all the crystals will swap direction. They still block the light, but they are all pointing the other way.

This makes driving an LCD a bit high maintenance, since you have to keep coming back to it to tell it to swap things up. It gets even more complex when

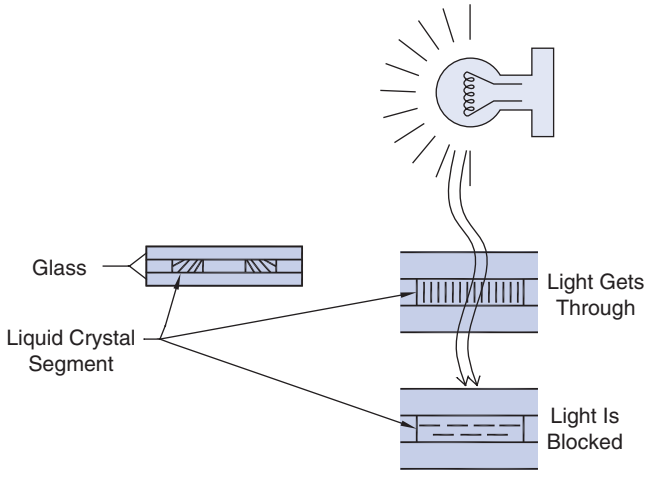


FIGURE 3.42
Inside an LCD.

you begin to multiplex the LCDs, too. You need to make sure you don't leave a cumulative DC bias on one of the segments too long, etc., etc., etc.[33]

For this reason, there are LCD driver chips. Sometimes this feature is built right into the micro; in other cases it will require a separate chip. You can go it alone and make your own driver, but I don't recommend it. It is easy to mess this up, and LCD drivers are pretty cheap.

Since it is an electric field that changes the LCD, driving the LCD is a bit like driving a capacitor. Every time you switch the LCD, a little current is used. Remember the RC circuit? But it is not much—in comparison to LEDs, it is virtually insignificant. You can get the current so low that a watch display can last for years on a battery. Remember, though, the larger the segment, the larger the cap,[34] and this means more current is needed to run the LCD.

Multiplexing

How do you do more than one thing at a time? Actually, you don't—you do several things quickly one at a time so that it appears that you are multitasking. (Like listening to your spouse while you are watching TV. A timely nod of the head can do wonders . . .)

---

[33]This isn't intended to be an exhaustive dissertation on the ins and outs of LCD displays. My hope is to simply get you enough knowledge to convince you to use a driver chip and save yourself a lot of headaches. It just isn't worth it. One chip I use extensively costs a mere 13 cents and handles 128 segments. I'll pay that dime any day!

[34]Remember that capacitance is a function of surface area.

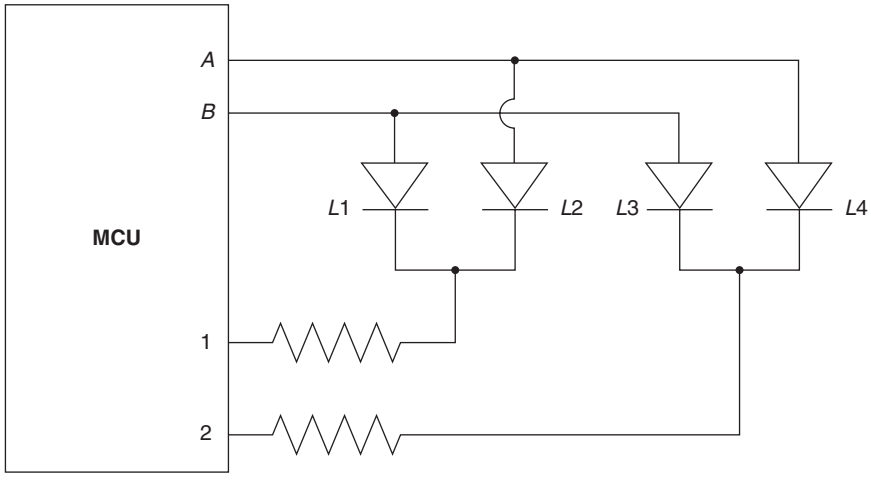


FIGURE 3.43
Multiplexing LEDs.

In the world of sparkies, it can be useful to multitask. One way to do this is by the art of *multiplexing*—that is, using fewer inputs to drive more outputs. Take a look at the example in Figure 3.43. In this case you can enable current to go through *L*1 and *L*2 by putting a low signal on pin 1 and a high signal on pin 2.

Due to the diode nature of the LED (think one-way valve) with a low on pin 1, putting a high on pin A or B will illuminate the appropriate LED. Reversing pins 1 and 2 will enable *L*3 and *L*4 to be illuminated. Repeat this process fast enough and to the human eye the LED will appear to be continuously lit. In this example we use four pins to talk to four LEDs, just to keep things simple, but increase the number of LEDs in each bank and you will quickly see how fast the number of LEDs you can talk to increases compared to the pins used. With three LEDs per bank, you have five pins running six lights; with four you have six pins running eight lights, and so on. If you have two banks of eight, you will have 10 lines controlling 16 LEDs! That is handy, especially when I/O is critical on that project where the PHB told you no, you can't have that more expensive micro with all the extra I/O. Remember, though, this slick application relies on the fact that the diodes pass current in only one direction.

Incandescent Lamp

Another indicator, the incandescent bulb is basically a light bulb. A resistive element in a vacuum tube heats up so much it gives off light. The fact that it heats up so much should trigger the light bulb over your head, saying to yourself,

"I bet that it uses a lot of current!" Which it does; it is rare that a micro has enough current capability to drive a lamp directly from a port pin.

Transistors and FETs

A bipolar junction transistor (BJT) or an FET is a great way to change the voltage (as we saw with the blue LED circuit) or to step up the output current capability of a micro. Don't forget to use a series resistor to the base with the BJT; you need to limit the current as you are switching a diode to ground. With the FET, protect the device from overvoltage or static shocks.

Coils

All sorts of devices have coils or inductors in them that you can send signals to. Let's take relays, for example. You might be able to drive them directly, but check the current requirements first! You will often need to use a transistor to handle the load. Also you will need a reverse-biased diode in parallel with the coil (to prevent excessive voltage spikes from causing damage). You can look at the section "Catching Flies" in Chapter 4 to learn more about the inductive kick on a coil and what to do about it.

Thumb Rules

- Use pull-up or pull-down resistors to assert an input signal when the input device is high impedance.
- Interrupt-driven inputs stop whatever the micro is doing while the line is active.
- Polling inputs allow you to control when you want to look at the inputs.
- Input devices come in an infinite variety of packages and capabilities, making the datasheet on the device very important.
- You can multiplex LEDs to scare up some needed I/O.
- Transistors are a great way to change voltage levels.
- Watch out for coils or inductors in devices; they will need some special consideration.

CHAPTER 4

The Real World

The real world is the place where you and I live. It isn't in this book or in a simulation or even the scribbles on a schematic. All those things are representations of the real world. They help us understand how the real world works. At some point, all the circuits we create and design will interface with the real world, even if it is just a button to press or display to look at. It follows that we should talk a bit about some of the things we use to hook our circuits up to the big, bad world.

BRIDGING THE GAP

If this book had been written back when computers were analog, this section wouldn't even be needed. As it is, the proliferation of those pesky little digital chips gives it top billing. You need to bridge the gap between the analog and the digital at some point if you want to market your latest gadget as "way cool digital technology." Knowing a bit about how to make the analog-to-digital leap seems like a good idea.

Analog vs. Digital

If we put analog in one corner of a boxing ring and we put digital in the other corner and then we let them duke it out, who do you think would win? In today's world, digital is all the rage, but what really sets it apart from analog? Let's find out.

What is analog? Is it merely some ancient term lost in the world of today's digital engineers? No; *analog* basically means *a continuously variable signal*. It means that the item being measured can be chopped up into infinite little pieces over time. Say, for example, a signal changes from A to B over a 1-second interval. If you look at it before 1 second is over it will be somewhere between A and B. It is a continuous variable. No matter how small you slice up the time segments, there is still a signal with information there.

The world as we perceive it is analog in nature. Colors blend infinitely from one end of the spectrum into the other. The sound as a car races by on the street is heard in a continuously increasing and then decreasing volume level. As you drive a car, you continuously change speed in response to the traffic and environment around you. The world around you is analog.

So what, then, is digital? "My computer is," you say. Yes, this is true. But let's get a little more basic with it. Hold up one of the digits on your hand. (A digit is your finger, in case you were wondering.) Now put it down, now put it up again. *This is digital*. It is either there, or it is not. I don't know if *digit* (as in finger[1]) is where the term *digital* came from, but it helps me remember what it means. So the simplest form of digital is two states: It's either there or not.

Let's go even deeper. What about the time it takes to change state? What if we look at our digital finger as it moves from all the way down to all the way up? If you look at it carefully, you see that a digital signal is really analog in nature. This is true. As one of my engineer friends is fond of saying, "There is no such thing as digital, really—just funny-lookin' analog." So digital is really just a mode of perception. You look at something in a specifically determined time frame and define whether it is there or it is not. Digital is a predetermined definition of analog levels.

If digital is really analog in disguise, why even bother with it? Early on it was discovered that digital signals worked well in communication. Remember the telegraph? It used a digital dot/dash series to represent a letter. Why does it work well? Let's look at our digital finger signal example again. At a distance, it is obvious to the observer whether your finger is up or down. In fact, this sort of signal is used on the freeway every day! All kidding aside, the point is that you can avoid communication errors by using digital signals for communication.

---

[1] Okay, so depending on the finger you pick for your personal example, you will either laugh or be offended. Either way, you don't want to let anyone see you give yourself the "digit" as you read this book in your cubicle. So I suggest the use of your index finger in this example.

So what is the drawback to using digital signals? The telegraph didn't last long. It was quickly replaced by analog forms of communication. The reason for this has to do with *bandwidth*, a measure of the amount of information a signal can carry. The analog signal can carry vast amounts of information. It can, in fact, have an infinite number of levels for a given signal range.

Back to the finger example: If you have a good telescope and can focus in on the finger from far away, you can easily see the varying levels that the finger can represent. The same thing can be accomplished without a telescope if you have a very large finger. This implies that analog signals can represent large amounts of information much more easily than digital signals can. To do this, though, you just need *more power*![2] (Imagine a manly grunt here.) If you can't get more power out of the signal, noise or other unwanted information can easily disrupt the signal. This is what happens when you get too far away from your favorite radio station and it starts to sound fuzzy. Sometimes you can give the receiver "more power" with better filters, amplifiers, or the like. Nevertheless, signal integrity is one of the struggles with analog systems.

On the other hand, to get a digital signal to move a lot of information, it has to work fast. Back when people wanted to hear each other talk, it was much easier to use analog signals. The digital technology of the time simply couldn't work fast enough to represent all the complexities of the audio information. Thus for many years communication efforts focused on analog encoding and decoding of information. However, digital was being used in another domain entirely, in the application of Boolean logic.[3]

Digital signals could be used to represent Boolean statements, one level indicating *true* and the other indicating *false*. The computer was born. Statements such as, "If *this* is true, then do *that*," could now be executed by machines. Boolean logic is based on a digital representation of the world. Don't think that there are only digital computers, though. For a while there were many analog computers in use to handle computations involving large amounts of information. Digital processing speeds eventually increased enough to take over these applications.

---

[2] I haven't met an engineer yet who doesn't like the idea of more power. Problem is, the need to meet the design specs (i.e., cost, design spec, weird management ideas, and so on) typically limits us in terms of the power that's available.

[3] If you like to immerse yourself in fascinating historical Internet research, I suggest you wiki or Google the name *Claude Shannon*. Shannon was considered the father of making circuits handle digital Boolean logic; his is an interesting story. Make sure you dig into his exploits in Vegas using information theory to take the house for a mint.

So We Have Analog

The upsides are that analog can represent lots of information, and the world around us can easily be represented by analog signals. The downsides are that it takes more power in either the transmitter or receiver to resolve the analog signals, and small analog signals can be easily disrupted by outside influences.

Then There Is Digital

The pros of digital are low power transmission and the ability to represent logic statements. The cons are information limits (low bandwidth), requiring it to work fast to process large pieces of information, and the fact that the world around us is analog, not digital in nature.

The Best of Both Worlds

Wouldn't it be great to have the best of both worlds? That's what engineers thought, so they coined a couple of acronyms to get the process started: ADC (the analog-to-digital converter) and DAC (the digital-to-analog converter). Let's find out what these are.

A-TO-D AND BACK AGAIN

What is A-to-D conversion (or ADC)? Is it a religious experience? Is it the opposite of D-to-A conversion (or DAC)? A to D is all about taking the real world and making it into ones and zeros so that digital technology can manipulate it. You can reasonably say that D to A reverses the process. Here we will explore what this A to D to A is and what it is good for.

A Is for *Analog*

An analog signal is converted to digital by chopping it up into chunks at predetermined time intervals. (This chopping is called the *sample rate*. The faster the sample rate, the higher the frequency that can be digitized.) Then the signal is measured at that point in time and assigned a digital value, which is called *sampling* the signal. Digital signals (typically represented as 1 or 0) can be crammed together to indicate different levels of analog. A single digit can indicate two levels. If you use a binary numbering system, the more digits you use, the number of levels goes up by 2 raised to the power of the number of digits. Four digits give you 16 levels (2^4). Eight digits gives you 256 levels (2^8) and so on. One common way of determining the level of a signal is to use a comparator, as shown in Figure 4.1.

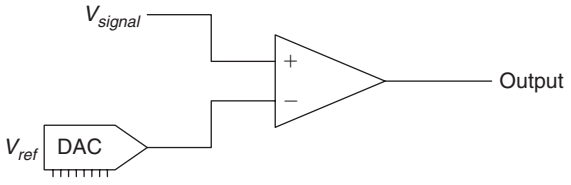


FIGURE 4.1
Comparator-driven ADC.

Study this application. In this case, the signal is compared to a reference voltage. You increase the reference voltage from min to max. When the signal is larger than the reference voltage, the op-amp comparator will output a high, or a 1. When the reference voltage is the larger of the two, the output will be low, or a 0. If the circuit knows the value of V_{ref} at the time the output changes state, this is when V_{ref} is approximately equal to V_{signal}. I say *approximately* because there is always a question of resolution. For more on this topic, read on.

D Is for *Digital*

Now that we have a digital signal we can do lots of fun things with it. We can transmit it, receive it, and manipulate it without worrying much about signal loss. But what is next? Say we convert guitar music into digital format to add some neat sound effects. You can't simply send the digital data back out to be heard. It must be converted back to analog. Why? Because there are certain things we perceive well in an analog format. If you don't believe me, take a look at the speedometer in your car; I'll bet it is an analog gauge. (There are some things we like to see digitally, but usually that's so we don't have to deal with infinite increments; look at the odometer in your car for this example.[4])

To convert a digital signal back to analog, the circuit has to simulate the analog signal it represents. This always requires some kind of filtering. There are many ways to convert digital to analog. One of my favorites is by *pulse width modulation* (PWM). In a PWM circuit, the device's output switches on and off at a given frequency—see Figure 4.2 on the next page. The percentage of time it is on versus off is the amount of analog signal it represents. This percentage is called the *duty cycle*.

The digital PWM is fed into a low-pass filter that removes the switching frequency of the signal, essentially leaving an analog signal. The number of levels that this signal can represent depends on the resolution of the PWM signal.

[4]That makes me think a bit. Is it human nature to prefer instantaneous signal references to be displayed in analog format whereas cumulative information is preferred in a digital format? Maybe some bright student out there will make this a thesis project so I don't have to think so hard about it. If you do, make sure you send me a copy; I'd love to know the results!

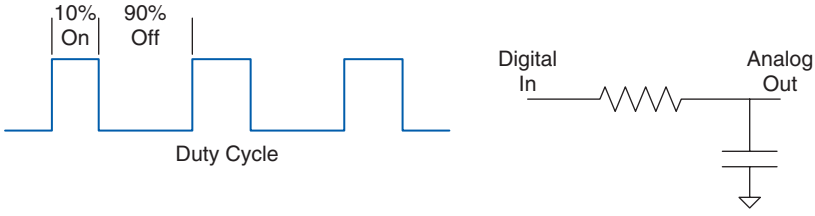


FIGURE 4.2
Duty cycle-controlled analog output.

This means the capability of the PWM to be switched on and off at varying duty cycles. For example, a PWM that could switch on and off in increments of 5% duty cycle would have less *resolution* than a circuit that can handle increments of 1% duty cycle—see Figure 4.2. This means that digital signals can represent only discrete levels of analog signal. These levels are the resolution of the signal.

Why is resolution so important? We stated earlier in the comparator example that the circuit knows what level V_{ref} is at. How does it know that? It must generate it somehow. It does so with some type of DAC process. It is the resolution of that DAC process that will determine the resolution of the ADC process.

So there we are. We went from analog to digital and right back to analog again. It really is a circle. Let's look at some examples to see this concept in action.

IT TAKES A LITTLE *D* TO *A* TO GET A LITTLE *A* TO *D*

A while ago I was explaining my thoughts on the world being analog in nature to a fellow engineer. He emailed me the following response:

> I would like to provide counterpoint to your assertion that "the world as we perceive it is analog in nature." I think that there are as many, if not more, natural digital perceptions as there are analog. Some samples: alive or dead, night or day, open or closed, wet or dry, flora or fauna, dominant or submissive, predator or prey, hungry or full, coarse or smooth, hot or cold, fuzzy or sharp, open or closed, single or multi, camouflage or warning, flat or mountainous, forest or desert, stormy or clear, noise or silence, blind or seeing, male or female, feast or famine, survive or die, on or off and so on. Granted, things like warm, breezy, sunsets, and omnivorous are there, but for the most part, I think our nature perceptions are digital.[5]

[5]This quote is from a good friend of mine by the name of Michael Angeli. I've always liked his writing style; maybe someday I can get him to collaborate on something with me.

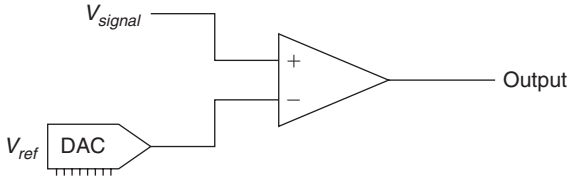


FIGURE 4.3
Comparator-driven ADC.

In many ways he is correct in his eloquent comments; however, he refers to our *perception*. We place the analog information from the world into "digital buckets." (There are certainly levels between hot and cold, for example.) I think the reason we do this is to facilitate decision making, to limit the store of information, and to ease communication. We impose a digital perception when it makes sense to do so. A better phrase I could have used is something like, *"The world is analog in nature, upon which we impose our digital perceptions."* With that in mind, let's look at some more of the nuts and bolts of A-to-D conversion. We'll start with the DAC and a simple comparator from a couple of pages back.

A simple comparator will output a high or low signal depending on whether one input was above or below the other (Figure 4.3). This is a great time to use a comparator, since digital circuits like obvious signals such as high and low. Let's drill the basic process of this circuit: You convert a digital number to a known analog level, compare that to an analog signal, and if it is close to the same value (here is where resolution counts), the digital number you output represents the analog value.

Let's do an example. You are converting an analog signal with the actual value of 4.45. You try outputting a 1 on your DAC. The comparator says "higher" (it does this by outputting a 1, or a high signal[6]). You then try outputting a 2. The comparator says "higher." Now you try a 3, then a 4. Guess what the comparator says each time. That's right, it says "higher." So what do you try next? Of course, you try a 5. Then the comparator says "lower." Now your circuit knows that the value is between 4 and 5. At this point you pick one of these two

[6]Remember that the specific voltage output of the comparator isn't important. At this point in the circuit you only care about the "state" of the signal. Is it high or low, 1 or 0, true or false? You get only those two options in a digital signal.

[7]It is important to note that you do not know to which value the actual signal is closer. You simply need to pick one. It really is an arbitrary decision and is fundamental to digital processing. This is the reason that resolution is so important. It narrows the gap and thus the lack of exact knowledge of the signal.

values[7] (assuming in this case that the DAC only outputs six discrete levels over a range from 0–5.) The smaller the steps or increments that you can output with the DAC, the closer you can estimate the value of the analog signal. When you make the steps in the DAC smaller you increase the resolution of the signal.

There is a better and faster way than merely sweeping across all the values in the range. (We will increase the resolution of our DAC now to illustrate this point.) Start by making your first output on the DAC equal to ½ of the entire range. In this case you output 2.5 on the DAC. Now look at what the comparator says and make a logic decision (digital is good for this sort of thing). You can see whether the comparator says "higher," and you can eliminate everything below 2.5. So you make your next output equal to half of the remaining range—in this case, you output 3.75. Look at the comparator again and eliminate some more possibilities (a high eliminates everything below the number, whereas a low eliminates everything above the number), then output half the remaining range. Repeat this process until you are out of resolution and you will have an approximation of the analog signal. This is a very fast way of converting an analog signal known as *successive approximation*. It is often used when high-speed analog-to-digital conversion is needed.

Did you notice that I often use the word *approximation* as the A-to-D process takes place? This is because a digital signal can never truly equal an analog signal; it must always draw the line somewhere. Do not forget that *digital* means that there are discrete steps involved. Analog has, by definition, infinite increments.

Now that you have the basic idea behind the A-to-D conversion process, let's look at some examples of DAC circuits to develop a more intuitive understanding.[8]

This is a slick way to get a digital voltage level, and you can get the R2R ladder in a nice compact package, as shown in Figure 4.4. You must take care not to hook it up to any low-impedance devices without buffering, since its output level can be easily affected by external loads.

How does the ladder work? A digital byte is output to the ladder, which changes the voltage level to the input of the comparator. You should note that the MSB (most significant bit) has the most effect on the output. The LSB

[8]More and more often these different types of DAC and ADC circuits are built into whatever part you are using. You might process a command that says, "Get me a sample of that signal." However, it is important to have an idea of what is going on in these parts if you want to be able to figure out why it isn't working the way you expected it to.

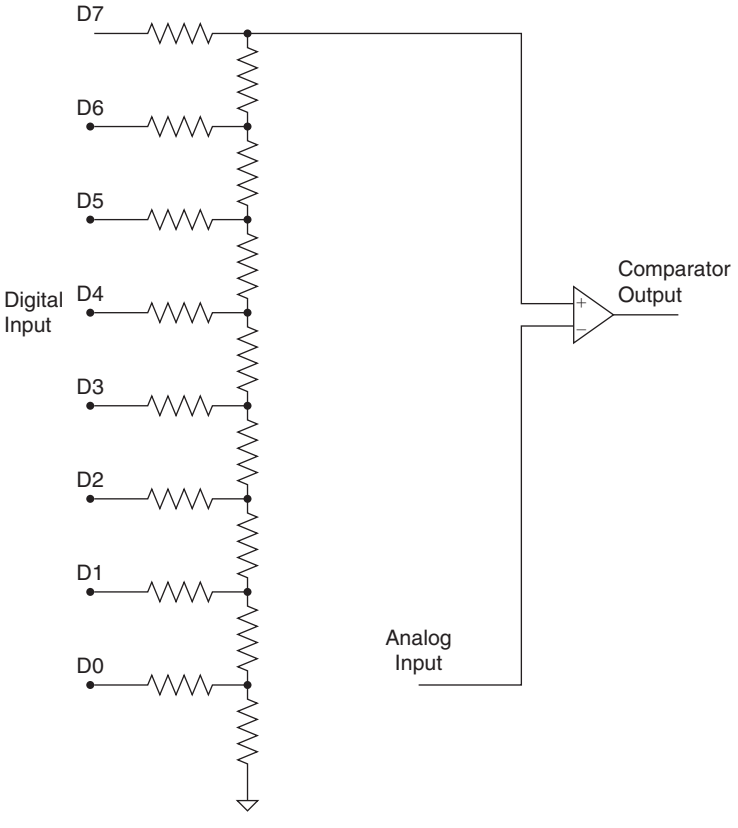


FIGURE 4.4
The R2R ladder.

(least significant bit) affects the output the least. This works very well with the approximation method described before. You simply load the DAC value you want on the resistors and look at the output signal. It is very fast. The biggest downside is that it uses a lot of output pins. (The output pins must be able to sink or source sufficient current to work correctly.) One caution: Make sure your processor can handle the output load of the ladder. The Zilog processor I used in one application of this circuit years ago worked fine and even had an onboard comparator for the ADC process, but I did use every pin, leaving little room for additional signals if needed.

In this circuit, the duty cycle of the PWM signal is ramped up from 0% till it passes the value of the analog signal, as indicated by the comparator—see Figure 4.5 on the next page. The analog voltage is represented by the percentage of the PWM signal when the comparator changes state. The RC filter must turn the PWM into basically an analog level. This means that the PWM must switch significantly faster than the speed of the signal you are trying to digitize.

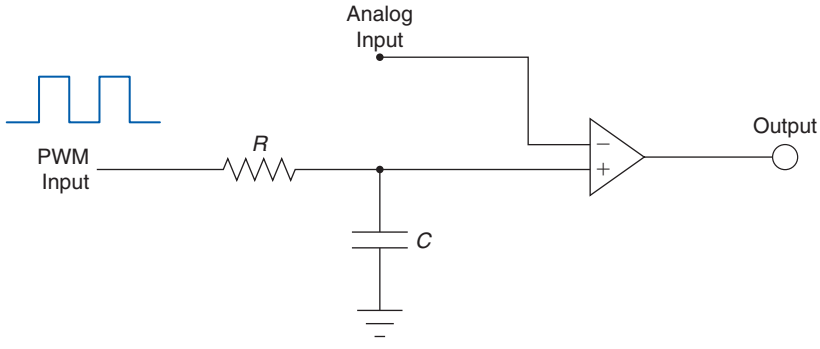


FIGURE 4.5
PWM ramp.

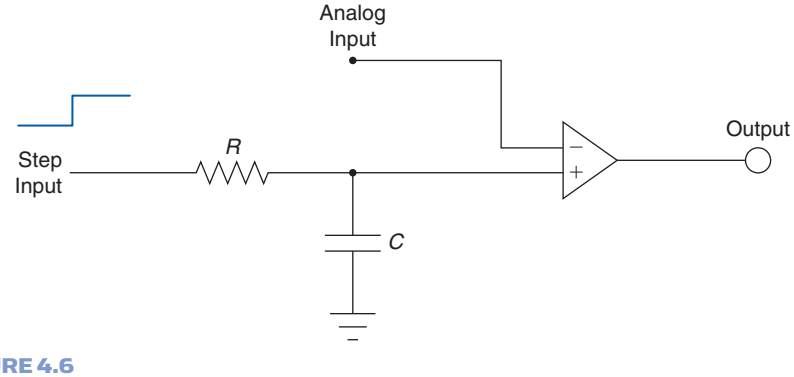


FIGURE 4.6
RC charge time.

This method relies on the transient response of the basic RC circuit (Figure 4.6). The step input causes the input to the comparator to increase according to the response time of the RC circuit. The output of the RC circuit is equal to $Vi(1 - e^{-(t/\tau)})$. So if you know the value of tau, which is R * C, you can calculate the voltage based on the time it takes to pass the input. This can be tedious to calculate in some micros, but often high accuracy is not needed and a lookup table of the values can be implemented. In many cases, a lower-resistance discharge path is added to this circuit to ensure that the output of the RC circuit begins at zero.

The downsides to this circuit are the curve calculations, but the first three tau of the signal are a fairly linear approximation. Depending on the application, that might be good enough. (Review the connection between electronics and hand grenades way back at the beginning of the book to see when things are "good enough.") If your task isn't too demanding and you don't get too close to the

upper rail, you can simply count time and toss out that complex calculation, making this a quick, cheap, and dirty ADC.

So there you have three easy ways to get a digital approximation of an analog signal. All these circuits are perfectly fine to use as DACs only.

One last thought: These days a built-in A/D converter is an increasingly common feature on a microcontroller. However, they nearly all work on the principle of using a DAC to make an ADC. Studying this section can help you to get an idea of what is really going on in there. The more you know about how it works on the inside, the better engineer you will be!

Digital Signal Processing

DSP, or *digital signal processing*, refers to manipulating data that is digitized from an analog signal. In many cases, such as audio and video, the signal is converted back to analog after DSP occurs. Many books are available on DSP that are far better texts on the subject than this one. Here I only hope to create a bit of understanding on this topic.

One of the advantages of a DSP is the ability to change parameters of the filters on the fly. This allows engineers to create all sorts of new solutions to processing signals that are very difficult to achieve with comparable analog designs.

Typically a DSP solution is also more expensive than an analog one, so be sure you really need it. Don't slap a five-dollar DSP chip in the circuit when a 25-cent op-amp will do the job. That is not to say DSP doesn't have its place. Without DSP, we wouldn't have MP3, WMA, AC3, AAC, MP4, WiFi, and a whole other slew of acronyms to spout about! Come to think of it, DSP technology might be responsible for more acronyms than any other . . .

Thumb Rules

- Analog is a continuously variable signal.
- Digital is a predetermined definition of a specific analog level.
- Digital signals have discrete steps.
- Resolution is the distance between the discrete steps.
- DAC is often used for ADC.

MAKING STUFF MOVE: THE ELECTROMECHANICAL WORLD

One thing that happens in the real world is moving stuff. Eliminating moving parts is a commonly sought-after goal in the world of electronics. However, I suspect that sometime in your career you will need to make things move and you will be thrust into the world of electromechanical devices. Considering that what I knew about motors when I left school could be written on the thin edge of a postage stamp,[9] I felt a need to cover some of the basics behind motors and a few other electromechanical devices here.

DC Motors

My eldest son was elated when he got a Lego Mindstorms kit for Christmas a few years back. For those who don't know, this is a ready-made robot kit based on—you guessed it—Legos. My wife claims I was much more excited than our son was. I beg to differ, but we won't go into that now. The whole point of a robot is that it moves. The Lego kit uses little DC permanent magnet motors with gears and such to get along. Since this type of motor is so popular, a little discussion about DC permanent magnet motors and how to control them seems prudent.

The DC permanent magnet (PM) brush motor is probably the easiest motor to understand. It consists of just a few parts: an armature, some magnets, a case, wires, and brushes. I remember as a kid making a motor out of a couple of nails, a dowel, and some wire. It looked something like what's in Figure 4.7.

You can make a motor by winding the wire onto the armature in a loop. The ends of the wire terminate on segments that the brushes rub on, as shown in Figure 4.8.

Permanent magnets are attached to the case in such a way as to surround the armature. The armature is supported in the case by bearings or bushings so that it can rotate freely. At its most basic, the coil of wire on the armature is nothing more than an inductor. As we learned earlier, an inductor develops a magnetic field when you pass current through it. This magnetic field is just like the one present around the permanent magnet. By controlling when the magnetic field is present around the armature, you cause the field around the

[9] My father was fond of this saying, so as a boy I spent more than a little bit of time looking at a stamp on edge and wondering just what you could fit there…

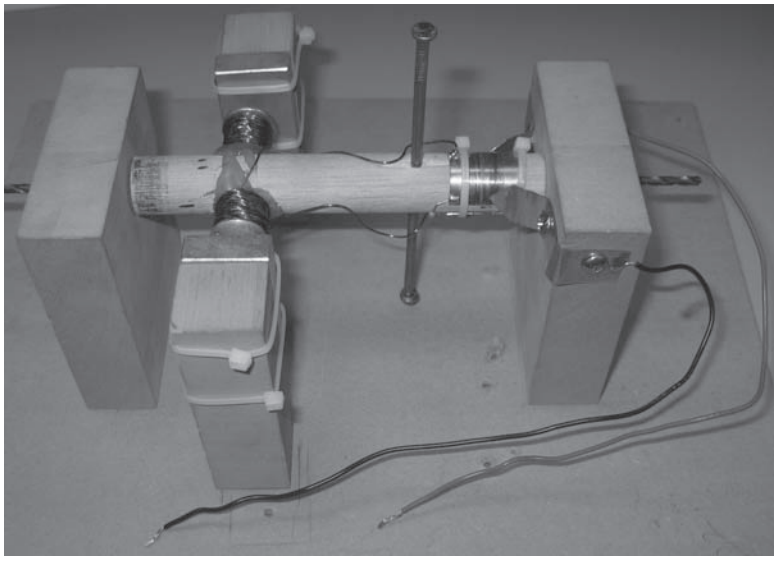


FIGURE 4.7
A home-built motor.

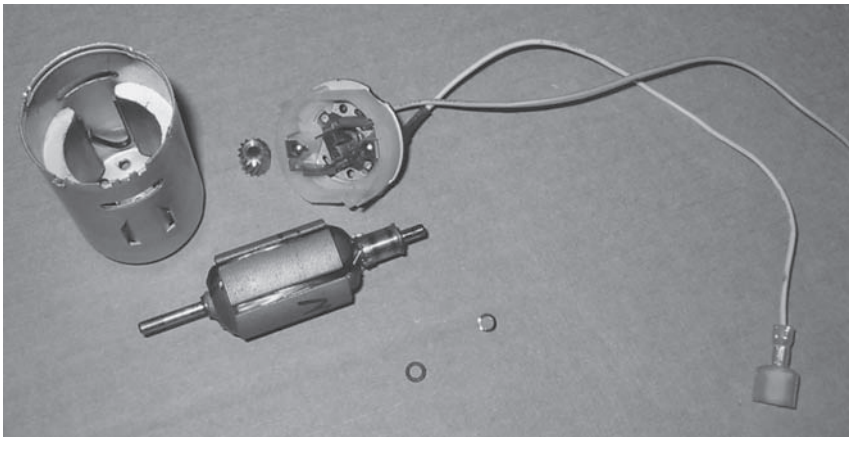


FIGURE 4.8
A motor taken apart.

wires to push or pull against the field around the magnet. The current to the armature is switched on and off (which turns the magnetic field on and off) in a sequence that causes the armature to turn. This is called *commutation*. In the DC PM brush motor, the brushes are the method of commutation. They switch the current through various sections of the armature as it turns.

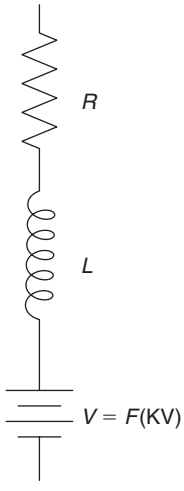


FIGURE 4.9
Inside a DC PM motor.

A DC PM motor has two inputs and two outputs. You put voltage and current in and get speed and torque out. One nice thing is that the speed is proportional to the voltage and the torque is proportional to the current. Motors are devices in which the physical equivalents of electric components are not only similar in nature but are actually linked in performance. Think of it this way: Voltage and current together equal power. Speed and torque together also equal power. So in a motor *you put electrical power in and get mechanical power out*. That actually makes sense, doesn't it? The equivalent circuit looks like the one shown in Figure 4.9.

What do you think the resistor is doing in this circuit? Have you ever noticed a motor getting warm when it operates? This heating comes from the resistive component in the motor. Any wire short of a superconductor has resistance. The armature, being made out of wire, also has resistance. Current flowing through a resistor will create a voltage drop across said resistor, and power across that resistor turns into heat. Ohm's Law still works.

The inductor creates the magnetic field that turns the armature. The battery represents what is called the back EMF, or *electromotive force*. If you were to spin the shaft of the motor with nothing but a voltmeter hooked up to it, you would see a voltage appear on this meter that is proportional to the speed at which you spin the shaft. When you apply a voltage to the motor, the shaft will spin at a speed in the same proportion. However, not all the voltage you apply to the leads makes it to this point in the motor. Some of it is lost across the resistor. All this leads to some characteristic equations of this type of motor.

The relationship between voltage and speed is known as the *voltage constant*, with units of volts per Krpm. It is referred to as K_e or K_v:

$$K_v = \frac{V - IR}{Krpm} \qquad \textbf{Eq. 4.1}$$

V = the amount of voltage applied at the leads

I = the current flowing through the motor

R = the equivalent resistance of the motor[10]

$Krpm$ = the speed of the shaft in thousands of revolutions per minute

The IR term in this equation accounts for the loss of heat in the motor. As current approaches zero, this effect disappears. This is what happened earlier when we hooked it up to a voltmeter and spun the shaft, reading the voltage generated. Do you see how that minimizes the error, giving you an accurate idea of the voltage constant?

The relationship between current and torque is known as the *torque constant*, usually referred to as K_i, which has the units inch-ounces per amp (in-oz/amp):

$$K_v = \frac{V - IR}{Krpm} \qquad \textbf{Eq. 4.2}$$

T = torque in inch-ounces

I = the current in amps

These two constants are linked; changing one changes the other. In fact, if you know one, you can calculate the other with this equation:

$$K_t = \frac{T}{I} \qquad \textbf{Eq. 4.3}$$

$K_t = K_v$	[Nm/A; V/rad/s]
$K_t = 9.5493 \times 10 - 3 \times K_v$	[Nm/A; V/Krpm]
$K_t = 1.3524 \times K_v$	[oz-in/A; V/Krpm]

As you can see, it turns out that we are really only dealing with one constant in the motor. This constant is controlled by the number of windings on the

[10]Note that you can get a fairly close idea of this with a simple ohmmeter turning the armature very, very slowly. (Too fast and the voltage generated will mess up the reading.) To be more precise, you need to take the resistance of the brushes and the way they contact the armature into account, a discussion that we will save for another book.

armature and the strength of the magnets. More windings increase the voltage/torque constant, fewer decrease it. The size of the armature and the strength of the magnets also affect this constant.

We now know that the main electrical components of a motor are resistance, inductance, and a voltage source. Can you extrapolate the mechanical properties? They are friction and inertia.[11] The first thing you should note is that the load, or whatever is hooked to the motor shaft, will likely be the largest contributor to these two characteristic factors, masking the effects of the armature inertia and brush or bearing friction.

Inertia will tend to make the motor take time spinning up to speed, increasing the load and current draw as you accelerate. Once at speed, inertia will tend to keep the motor spinning, so during deceleration you will notice a lessening of the current the motor needs.

Friction will create a constant load on the motor that will appear as an increase in current in our "sparky" universe. To gain further light and knowledge on all things motor, I refer the reader to the "pink book."[12]

"But what about the Lego Mindstorms?" you ask. Well, my boys and I have built quite a few projects, but my oldest son lost interest since we couldn't seem to build a robot that would clean his room. I told him that I sincerely hope he can someday solve that particular problem, but till then it is up to him to get the job done.

DC MOTOR CONTROL

Given what we just learned about this type of motor, I hope it is apparent that if we want to control the speed of a DC permanent magnet motor, we should control the voltage to it. If we want to control the torque, we should control the current. If this doesn't make sense, take a look at the DC motor equations once more.

SPEED CONTROL

Let's start with a simple application. Say you want to spin a motor at 500 rpm. This motor has a K_v of 10 V/Krpm. Plug that into the equations we just learned

[11] You could also have a spring-type component, as we discussed way back in the beginning of the book, but it is pretty rare to find that in a DC PM motor.

[12] *DC Motors Speed Controls and Servo Systems.* I like to call it the "pink motor book" due to an interesting choice of color for the cover. I highly recommend it for anyone who is working with DC motors.

and you find out you need about 5 V to get this motor going at the speed you want (neglecting load for a moment). But how do you go about getting 5 V to the motor? There are two different ways to approach this problem: You can use a linear methodology or a switching methodology. In both cases, you will start with a higher voltage than you want at the motor and then lower it and apply it to the motor leads. We should go over both types of systems to understand the pros and cons of each.

LINEAR CONTROL

The simplest way to make a linear control is based on the voltage divider rule. Put a resistor between the power supply and the motor, and adjust the resistor's value until you have the amount of voltage you want across the motor—see Figure 4.10.

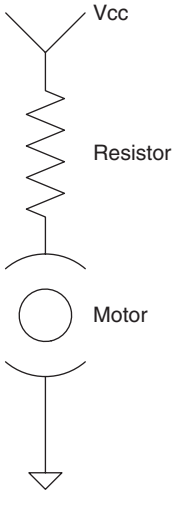


FIGURE 4.10
A motor with a resistor in series.

The biggest drawback about this design is that the motor can be a dynamic load. As the load on the motor changes, the amount of current through the motor changes, which, following Ohm's Law, changes the voltage drop across the resistor, which changes the voltage across the motor and hence the speed of the motor is destined to change.

However, if the load is consistent or if variability is okay, you can dial this design in and make it work fine. You should note that the resistor will heat up based on the current through it and the voltage across it. For example, if *Vcc* is 10 V and you set the resistor value such that 5 V is across the motor, this

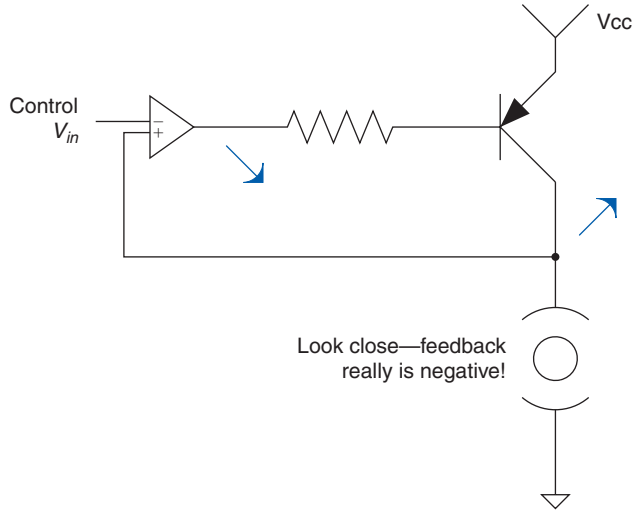


FIGURE 4.11
Op-amp-controlled motor.

means that there is 5 V across the resistor. If the current drawn by the motor in this case is 1 A, you will need a 5 W resistor to handle the power. (Actually, any engineer worth his salt will not run the power resistor at its maximum wattage but will overate it liberally.)

In this linear control design, the resistor can be replaced by a FET or transistor or some other type of amplifier operating in linear mode, allowing the voltage to the motor to be adjusted as desired. By using feedback methods as previously learned, the variation in load can also be compensated for so that you can maintain the desired voltage to the output. Look at Figure 4.11.

This design has the significant advantage over the previous one of maintaining the voltage to the motor at a desired level, regardless of the changes in load. This will maintain a more constant speed than the previous design, but there is still room for improvement, as we will see later.

The biggest drawback to this type of design is the same as the resistor. Excess power is turned into heat. One benefit, though, is that as far as EMI is concerned, it is a quiet design.

SWITCHING CONTROL

In contrast to linear control, a motor can also be controlled by switching power on and off to the motor. The similarities of switching motor control to

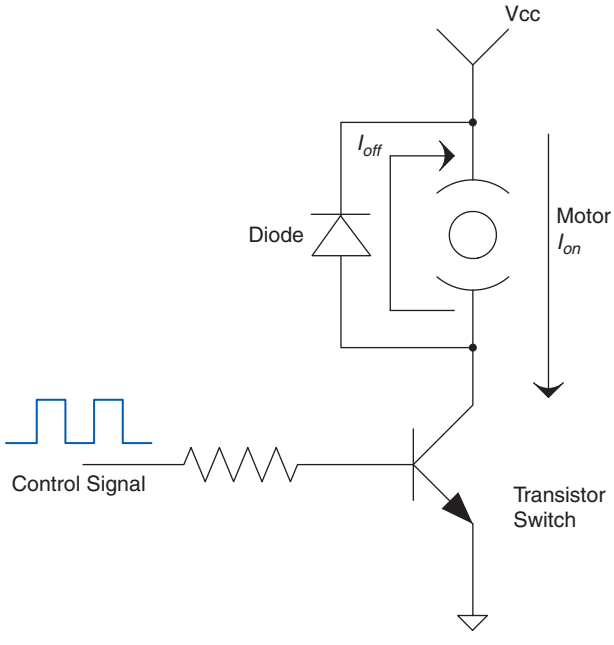


FIGURE 4.12
Transistor with motor and diode.

switching power supplies are many. In many switching supply designs you will find an inductor that stores the energy when the switch is on and discharges it to the load when the switch is off. The same thing can happen in a switching motor control. The inductor, however, is inside the motor. In the switching supply you will find a diode that directs the current from this inductor to the load. In a correctly designed switching motor control, you will find a diode that performs exactly this function, as shown in Figure 4.12.

When the switch is closed, current flows through the motor. When the switch opens, the current from the inductor goes through the freewheel diode back around through the motor. Since this current is recaptured and applied to the load, switching motor controls are more efficient than their linear counterparts.

Some important things to note: The switching frequency of this type of control needs to be fast enough for the inductor in the motor to act this way. If the switch doesn't turn back on before all the current has discharged from the inductor, you will feel the torque change in the motor (it will manifest itself as a vibration).

In a way, the inductor filters the high frequency of the switching power, reducing torque ripple and thus vibration. The most common form of control in this case is called *PWM* for *pulse width modulation*. By varying the duty cycle of the PWM, the amount of power to the motor is varied.

Switching motor controls are very prevalent these days. This is primarily due to their efficiency and the proliferation of high-power, high-frequency switching devices.

MAINTAINING SPEED

Often some sort of voltage feedback is used to maintain the output voltage of the control to a desired level. Remember that in the DC PM motor the speed of the output shaft is proportional to the voltage applied. That makes it nice for maintaining speed. However, if you flip back a few pages you will notice the *IR* component of that equation represents what are known as *losses*. These losses are burned up as heat across the resistance of the wire in the motor armature (plus a little in the brushes). The loss is proportional to the current (*I*) through the motor, and the current is proportional to the load on the motor shaft.

This means that as load varies on the motor, the amount of loss varies. This results in a change of speed. Think of it like this: The voltage that gets burned up as heat never makes it to spinning the motor shaft—see Figure 4.13.

There are two ways to compensate for this. One way is to use speed feedback to adjust the voltage output to the motor to maintain a constant speed. The other is to compensate for the losses themselves.

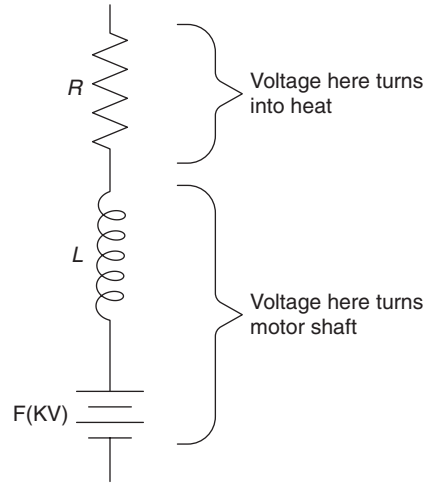


FIGURE 4.13
What happens inside the motor?

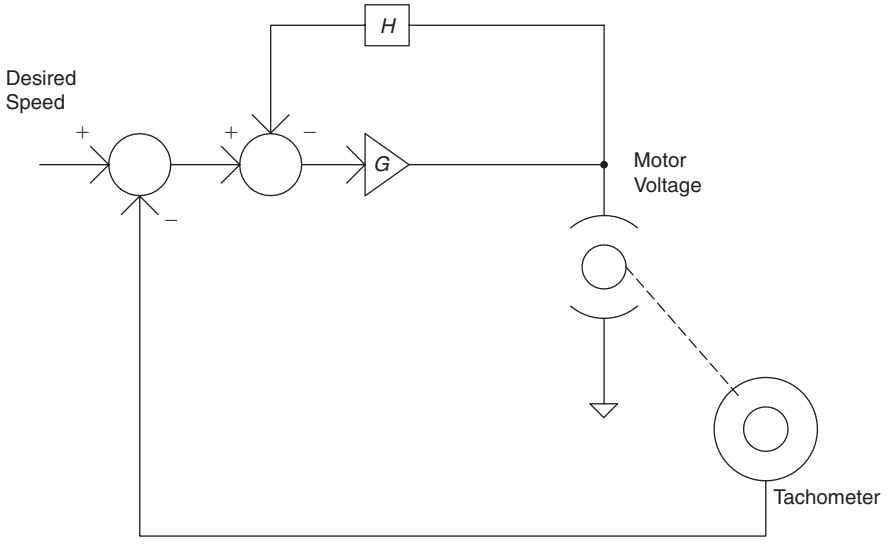


FIGURE 4.14
Block diagram of speed feedback.

In most DC motor control designs, you will find a voltage feedback loop that does 90% of the speed control work. Then, external to that, you will find a speed feedback control loop that will compensate for the rest of the variation—see Figure 4.14.

Though it is generally a good idea to feed back the signal you want to control, there could be times you do not have that luxury, or maybe there are reasons you do not want to use speed feedback. If this is the case, you can use another speed control approximation called *IR compensation*—see Figure 4.15 on next page. This is a method in which you monitor the load on the motor by sensing the current through the motor. The loss due to heat is proportional to this current. If you know the resistance of the motor, you can calculate how much voltage turns into heat, never making it to the output shaft. Add this much voltage to the input to the motor and you have a fairly good approximate speed control and you didn't need a tachometer!

All in all, controlling a DC PM motor speed is one of the simplest motion control problems to tackle, but it is simple only relative to the other options out there. It is still a significant "chunk o' learnin'" to swallow. For this reason I suggest starting here if you want to learn motion control before moving on to some of the other available motors.

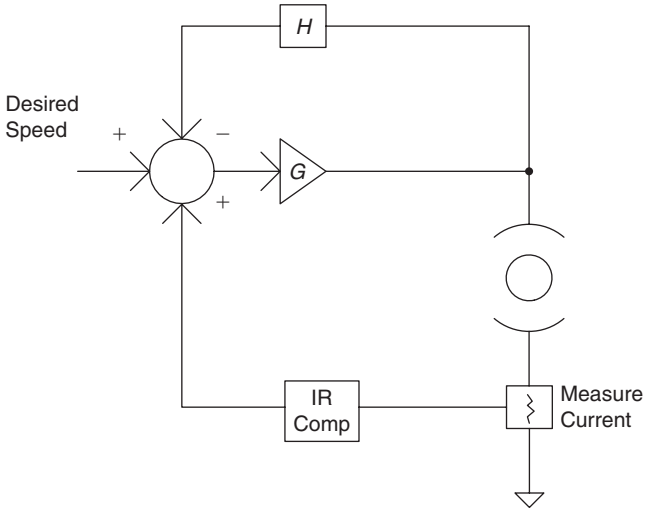


FIGURE 4.15
Block diagram of IR comp feedback.

Torque Control

One thing that happens when you take a motor and stop the rotor is a huge increase in current. Depending on the motor and your design, this could be more current than your control can handle. In this case you might need a current-limiting circuit.

This is a circuit that monitors the current used by the motor and, at a preset level, scales back the output, limiting the current available to the motor.

In a DC PM motor, when you control the current, you control the torque. Don't believe me? Flip back a few pages and look at the torque constant equations. The units are in oz per amp. This is a linear relationship—the more current through the motor, the more torque at the output shaft.

What all this means is that a constant current supply will create a constant torque when hooked up to a motor. This is essentially what happens when a control hits its current limit. The control goes from being a constant voltage supply to a constant current supply. This protects the motor and the control from damage.

Braking

Imagine careening down a hill on your electric scooter. "Gosh," you think to yourself, "it would be nice to use some of the energy I'm wasting to slow this vehicle down. There ought to be a way to make it recharge the batteries.

Hey, I'm an engineer," you say to yourself, "Why don't I design a regenerative brake?" Just such a thought has come into my head and I have been able to ignore[13] it quite effectively until now.

Some time ago I was asked to design a motor control with a regenerative braking circuit. Having done several controls, but none with regenerative braking, I started by perusing the Internet. I don't follow *Star Trek's* creed to boldly go where no man has gone before on a whim. That is to say, if someone has been there already, I would sure like to know the path he or she took. Once the end of that path has been found, I will then venture into the unknown.

Anyway, in this case, several hours of searching were somewhat futile. A simple and concise explanation and possibly a schematic (particularly for a PM DC motor) were all I needed. There were reams[14] of information explaining what it does but not much was there showing exactly how it was done. Alas, my effort to find the simple explanation was to no avail. Maybe it was out there somewhere, but I got sick of all the pop-ups.

As you might have guessed by now, I take such a lack as a personal affront that I must correct. The following is what I have pieced together in my own mind, distilled down to my level of intelligence (the longer I spend in management, the lower this level seems to be), then ousted to my readers in a form I hope is easy to understand. After I looked at the best idea since raw toast and the nice read about the Honda Insight's regenerative brake, the following is what came out.

No More Secrets!

One place I found said that regenerative braking is the well-kept secret of motor control. However, when I learned the truth, I think it is just poorly explained. Let's start with Figure 4.16 (see next page), a diagram of a simple PWM controller for a DC PM motor.

A PWM is fed into a switch, such as a MOSFET, at a frequency that is high enough to keep current flowing in the inductor inside the motor, not at all unlike a switching power supply. When the PWM shuts off, the current flows through the diode (sometimes referred to as a *freewheeler diode*). That part I could understand, but the question that I kept asking myself was how do you

[13] I find it very easy to ignore such thoughts when I am playing Nintendo (or Xbox 360). In fact, back in my college days, I had to redo an entire quarter of school due to a severe Nintendo addiction (except for one class that I passed due to a very persuasive paper on said topic). But we'll save that story for some other time.

[14] Can you use the word *reams* when referring to the Internet? After all, it isn't really on paper, is it?

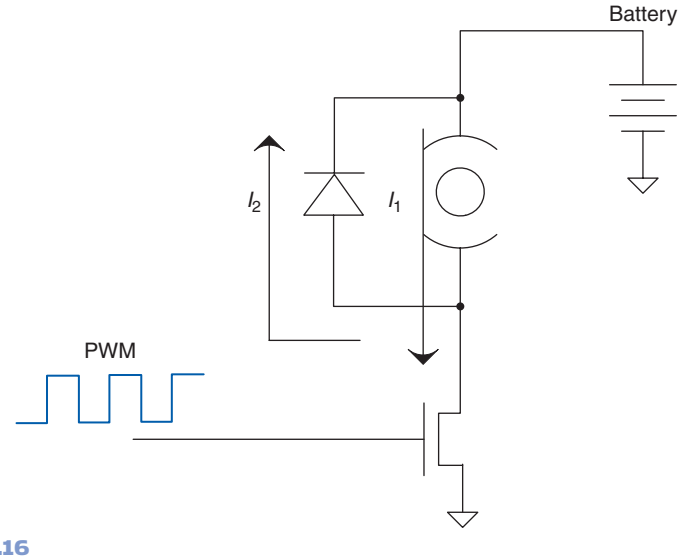


FIGURE 4.16
PWM motor control.

get a motor, which is spinning at a lower voltage than the output of the battery, to push current back into the battery?

Let's start with a small change to our earlier circuit, as shown in Figure 4.17. We will replace the diode with a synchronous switch that goes off when the primary

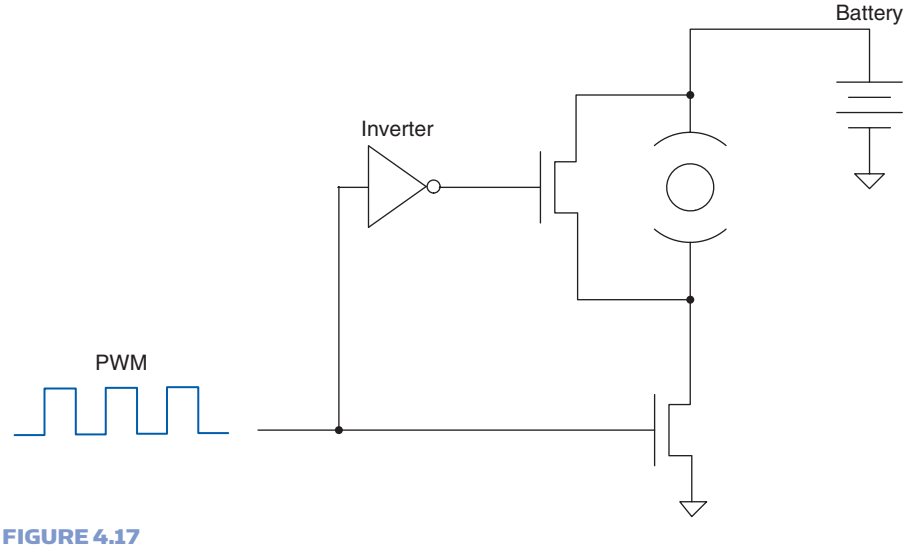


FIGURE 4.17
PWM motor control with FET in place of diode.

goes on, and vice versa. For the purpose of this discussion we will ignore the fact that the FETs need particular driving methodologies for the high side and the low side of a motor.

I had read about this topology many times. It is usually brought up as a way to make your controller more efficient in terms of heat loss. This is because the FET has a significantly lower voltage drop across it than the diode does. I had no idea that it also functions as a regenerative brake, until I figured it out for myself. Here is how it works.

A Little Elaboration

Keep in mind that in the original version of this controller there is a voltage across the motor that depends on the duty cycle of the PWM, but it is referenced to the positive output of the battery, not the negative side. That helped me to keep it in perspective.

Assume that we have a 12V DC battery and there is 6V DC across the motor. That means you would see an average of 6V DC from the bottom of the motor to ground. Now let's say that you spin the motor faster than 6V—for example, 7V. If you keep the same average voltage at the bottom of the motor, you will have 1V extra to dump into the battery. This explanation doesn't entirely jive, but I think it will get you in the right frame of thinking. If you follow it to its conclusion, you will think that the previous version with the diode should also regenerate, but it does not.

Let me elaborate. With the diode version, there is no braking force generated. That comes into play when the diode is replaced by the FET. When the freewheel FET turns on, the voltage generated by the motor is shorted back into itself. This provides the braking force and a current flow in the opposite direction through the motor. Remember the rule of inductors (since there is a decent-sized inductor in the motor). Once a current is flowing, it doesn't like to stop. So when the high side opens and the low side closes, current is pushed into the battery. Wave of the wand and voilà, you have regeneration!

Regeneration Ain't So Bad

It turns out that regeneration isn't so tough at all. In fact, it is almost a side benefit of making your controller more efficient, if you want to look at it that way. Now if there were just some way of making it more than 100% efficient, hmm . . .

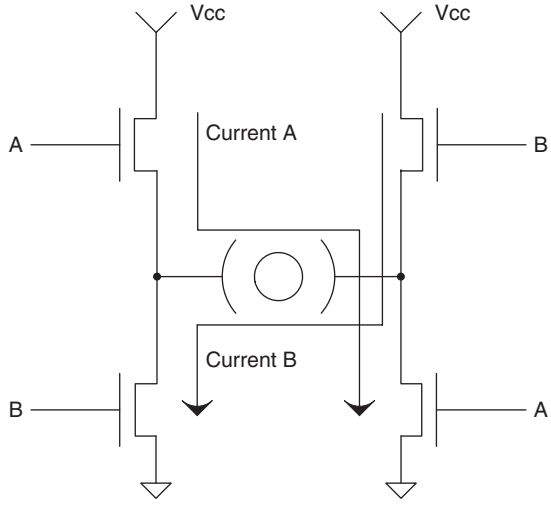


FIGURE 4.18
H bridge motor control.

Changing Directions

In a DC PM motor, it is fairly easy to change the direction of spin of the armature. You simply need to reverse the voltage to the motor leads. A common way to do this is known as an *H bridge*, for the way it looks when drawn on a piece of paper, as shown in Figure 4.18.

As previously discussed, the H bridge can be a linear or switch mode design. The same theory is applicable, but it does become more complex. For example, you don't want to turn on both legs on one side of the bridge because you will create a short across your power supply. This is known as *shoot through* and will usually cause copious emissions of magic smoke.[15]

Synchronous switching of the opposing high and low legs with the appropriate devices (like FETs) will create the braking/regenerative effect that we have already mentioned. Voltage and current feedback become more complex due to the fact that one leg of the motor is no longer tied to one spot all the time. You will need some differential amplifiers that can handle some large voltage swings to get things working right.

[15] I have found that letting the magic smoke out of a component can be very entertaining. But once the magic smoke is gone, those parts just aren't the same anymore.

Making Stuff Move Conclusion

Controlling motors is one of the most complex and rewarding things you will do in electrical engineering. Setback and frustration will be rewarded with the pride of seeing something move! There is no way I can possibly cover all aspects of motor control in this text. I do hope, however, that I have given you enough basic understanding that when you tackle motors and do more research on the topic, you will be able to understand what you find out there.

SOME OTHER TYPES OF MOTORS

You will run into many types of motors. People have been goofing around with different ways to make a motor nearly as long as they have been messing with electricity. Here is a bit of overview on some various types of DC motors.

Brushless DC Motors

The brushless DC motors shown in Figure 4.19 are cousins to the DC PM motor we discussed earlier, but instead of using brushes for commutation, they usually use some type of electronic control. To accomplish this, usually the inside of the motor has the permanent magnets (where the armature is in the DC PM motor). This is known as the *rotor*. The windings are on the outside and are usually referred to as the *stator*, or field windings.[16] There is no requirement for

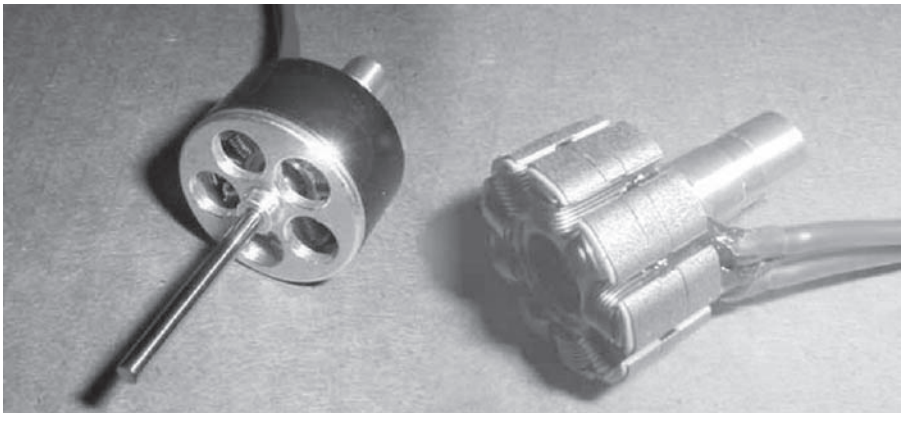


FIGURE 4.19
Cool little brushless DC motor out of my RC airplane.

[16]You can flip-flop the magnets and the windings. You might see a motor with the magnets on the outside and the windings in the center. The windings are still stationary and the magnets still move. The RC world calls them outrunners.

the magnets to be on the inside and the windings on the outside, but the windings are stationary and the magnets rotate. The rotor is turned by switching the stator windings on and off in a sequence that creates torque on the rotor. This is known as *electronic commutation*, as opposed to brush commutation, which we already learned about.

Often you will be told how a DC brushless motor is so much more efficient than a DC PM brush motor. There is some hype to dig through here. Though this claim can be true, it is not entirely due to the fact that it is a brushless motor, as the brushless motor guys would have you believe. You will see numbers showing improved efficiency, but that is generally due to the choice of magnets. Most brushless motors are using rare earth magnets that have a much higher flux density than the more common ceramic kind. What this results in is fewer turns of wire for the same torque and speed. Fewer turns of wire means shorter wire, which means lower resistance. Since the resistance of the windings is the largest loss in the motor, this makes the motor more efficient.

DC PM motors commonly use the ceramic magnets, resulting in more turns of wire. To make them more efficient, you need to increase the wire diameter to lower its resistance. It is possible to use the stronger magnets in a DC brush motor. The most common place I have seen this is in hobby stores. There are some pretty cool motors like this for RC airplanes. When built with these "super" magnets, the DC PM motor is pretty close to the same efficiency as the DC brushless motor.

Assuming good bearings, the next point of loss in a motor is in commutation. In the DC PM motor the brushes and brush contacts are the method of commutation. This interface is not perfect and creates a resistive loss. In the DC brushless motor, commutation is done with some type of silicon switch such as an FET, for example. Typically it takes at least six of these parts to commutate a DC brushless motor. These FETs have a resistive component (RDS on) that causes loss in the form of heat.

The biggest advantage to a brushless motor is right there in its name. It has no brushes. The brush in the DC PM motor will nearly always be the first thing to wear out. Brushes by their nature are designed to wear out, but don't let that stop you. There are many types of brush motors available, and often they will be just fine for the application.

One thing to note about brushless motors is that the controllers are more complex, requiring three to six times the power devices that brush motors use. But once you have them under control, you have already spent most of the money

needed to make them go in both directions. So if that is a feature needed, it could make a brushless motor more of a candidate.

Stepper Motors

Stepper motors are a type of DC motor in which the output moves a specific distance each time you energize a winding. They are a cousin to the brushless motor and a weird animal called the switched reluctance motor.[17] The ability to move a specific step makes these devices commonly used in positioning mechanisms. Printers use them by the bucket load.

Positioning is relatively easy since you can energize the windings and count the number of steps you have made to determine where the motor shaft is.

Stepper motors are characterized by their moving torque and holding torque. This is important to know because if you exceed either, your motor could slip, and that would cause your count to be off.

Thumb Rules

- In a motor, you put electrical power in and get mechanical power out.
- Voltage * current = power; speed * torque = power.
- Linear controls cause less EMI.
- Linear controls are simple and cheap.
- Linear controls are less efficient due to heat loss.
- Switching controls are more efficient.
- Switching controls cause more EMI.
- Switching controls are generally more complex and expensive.
- Constant voltage makes for constant speed with a DC PM motor.
- Constant current makes for constant torque with a DC PM motor.
- Don't forget the freewheel diode in a switcher.
- Replace the freewheel diode with an FET and you have a brake.
- Use an H bridge to change directions.
- Brushless motor controls are inherently bidirectional.
- Stepper motors move in small steps or increments.

[17] Somewhere between an AC motor and a brushless DC PM motor lies the switched reluctance design. It is rare enough that the reader is left to his or her own resources to find out how this unique design works.

AC and Universal Motors

As we mentioned earlier, long ago a smart guy by the name of Tesla helped us all by convincing the powers that be that we should have an AC means of power distribution (vs. the local DC generators that Edison wanted). One key factor that helped with this debate was Tesla's invention of the AC motor.

There are many types of AC motors. One of the most common and the one we are going to review here is the *AC induction motor*.

An AC induction motor induces a current in the armature by varying the magnetic field in the stator. This induced current in turn creates a magnetic field that causes the rotor to turn, pushing against the first magnetic field. When I first learned this, it seemed to me that an AC motor can pick itself up by its bootstraps, so to speak. One result is that the motor tends to have a "sweet spot" where the rotational speed is just right, generating maximum speed and torque. At lower speeds the torque drops off pretty fast. This leads to the fact that AC motors are not known for low-speed torque (unlike the DC versions we just discussed). For this reason and the fact that AC motors run off a sinusoidal alternating signal, a huge percentage of AC motors are fixed-speed outputs where the speed depends on the frequency of the AC signal. There are variable frequency drives or controls, similar in architecture to DC brushless drives. They can vary the frequency into an AC motor, creating a variable speed AC drive. Since AC motors do not have such a simple torque speed curve, these controls can be fairly complex, often using DSP chips to handle all the math needed to get what you want out of one of them.

AC motors have been around for years, making them relatively inexpensive, and their lack of brushes makes them last a long time. They can be built synchronously, like a stepper motor, so that you know they have moved a set distance every cycle of the AC wave. You will see them in all sorts of places: running compressors in a refrigerator to timing the icemaker circuit in the same fridge. Back before the "day of the diode," they were used in millions of clocks.

Universal motors are like PM motors without the permanent magnet. They use windings with current flowing through them in the outer field instead of said magnets. What makes them universal is the ability to wire them to work with an AC or DC source. I shocked myself more than once rewiring the motor down in the old milking barn trying to figure this out.

Motors of all shapes, sizes, types, and voltage preferences are out there. Hopefully I have provided enough background so you at least sound smart when you're asked about this topic.

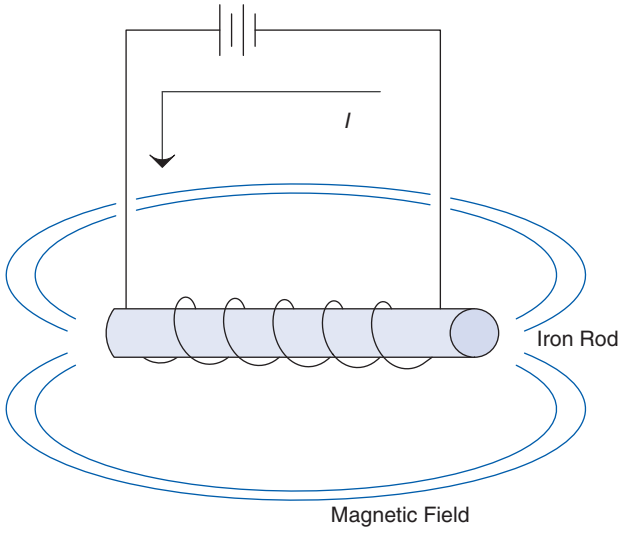


FIGURE 4.20
Iron rod in coil of wire.

Solenoids

The *solenoid* is an electromagnetic device that typically moves to only two positions. Akin to the stepper motor, solenoids are rated by holding force and moving force.

Take a coil of wire and an iron rod that just fits inside the coil, as shown in Figure 4.20.

Energize the coil. The rod will center itself in the coil due to the magnetic flux running through it. It is in fact reluctant to leave the warm home of its cozy little coil. This tendency for ferrous material to align itself with magnetic fields is known as *reluctance*.

Shut the magnetic field down and the rod moves easily. Usually a spring is used to push the rod out of its cozy coil shell until power is switched back on and the rod returns to showing its reluctance yet again.

Solenoids are great if you need a short linear motion that is controlled by something electronic. Also, this concept is the basis for an electromagnetic cannon. We don't have time to cover that topic in this book, though. Too bad; cannons are fun.

Relays

Relays don't actually make anything move except the part inside them that closes a switch, so they might seem a little out of place in this discussion. However, they are definitely electromechanical in nature and I couldn't think of a better place to talk about them. Relays are very tough; they predate the transistor by a long time and are still in use. That should say something. They are basically the combination of a solenoid and a switch. A magnetic force pulls the switch shut or open, depending on the particular device. Markings on a relay usually indicate the coil that operates the relay and the labels NO, NC, and C. These abbreviations, also sometimes seen on switches, mean normally open, normally closed, and common, respectively. NO and NC refer to the state of the switch when the coil isn't energized. C is a connection to both these switches.

There are two important specs on a relay: the coil voltage and the contact ratings. If you under-drive the coil, you might get the switch to close, but there are no guarantees. Contacts are often rated at a minimum as well as a maximum current. Most engineers are diligent about paying attention to the max current, but they often ignore the minimum current. Many relays used in a power setting (which is very common these days) rely on a certain amount of current to be present when the switch opens. This current creates an arc that cleans the contacts and keeps them from corroding. Do you have a relay that simply stops working after a while? Chances are you are not meeting this spec. Use a relay in your design and you get to hear that satisfying click, letting you know that something is really working in that magic box.

Catching Flies

One thing all these motors, solenoids, and relays have in common is a coil of wire that is switching current at some point. A coil of wire is an inductor, and an inductor doesn't like current changes.[18] So what happens when you shut off the current in an inductor? As the magnetic field collapses when you cut off the current, a large voltage spike is generated (because it wants to keep current flowing). This spike is sometimes called the *flyback*. To keep this spike from damaging components and to use the energy in it, most applications employ a flyback or, as it is also called, a *freewheel diode* that shunts this spike back to its source, as shown in Figure 4.21.

[18] By now this phrase should feel natural and intuitive. If it doesn't, go back and study an inductor and how it relates to current till this concept makes sense.

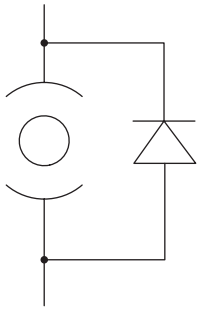


FIGURE 4.21
An example of a flyback diode around a motor.

If the response of the LR in this circuit is slower than the switching frequency, the diode acts as part of a filter keeping current moving through the motor. It smoothes out current changes, which in turn smoothes the torque changes. (Remember how torque is proportional to current?)

In other cases, this diode might simply be capturing a transient signal to prevent circuit damage. Figure 4.22 shows an example using a diode to protect a relay circuit.

You can see that the voltage spike, inductive kick, or back EMF, as it is called, never gets over –0.7 V because once it does, it forward-biases the diode and

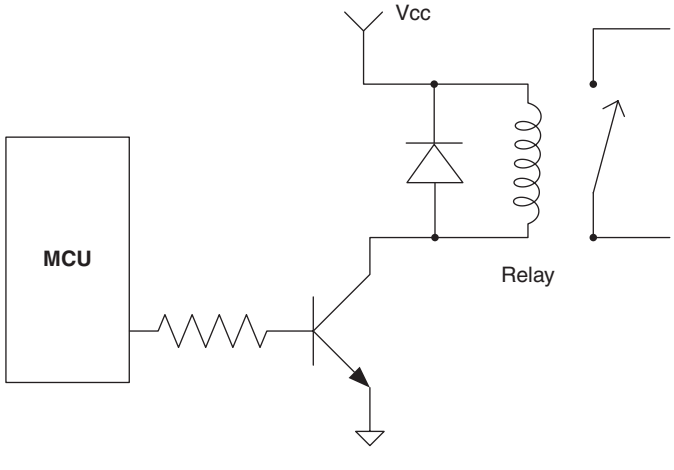


FIGURE 4.22
Flyback diode on a relay coil.

current flows back into the other end of the inductor. Now you know how to make a fly catcher out of a diode.

More Thumb Rules

- When the thumb rules go on and on, break them into smaller, more digestible pieces.
- AC induction motors induce a current in the core, which in turn creates a magnetic field that turns the shaft.
- Universal motors can be wired for AC or DC.
- Solenoids are reluctant to leave the cozy coil cave when current is on.
- Pay attention to the minimum switching current on relays.
- Catch your flies with diodes to keep voltage spikes out of your circuits unless you are trying to make a shock box to surprise your buddy.

POWER SUPPLIES

Whatever you do with electronics, you are going to need power to accomplish it. It will be useful to understand the basics of power supplies, since you are nearly guaranteed to deal with them at some point in your career.

It's All About the Voltage, Baby!

Most devices today want to keep the voltage constant. This means that current can vary as needed. In the world of power, particularly as it relates to the ubiquitous IC, it often seems that you never have the exact voltage you want.

A huge number of products run off 120 V AC out of a wall socket. Another huge group runs off batteries that are charged from those wall sockets, and another significant number runs off batteries that you can buy by the caseload at any super-duper-mart. Just ask yourself, how many batteries did you buy last Christmas?

The problem is that most ICs these days want 5, 3.3, or even 1.5 V DC. This is nowhere near 120 V, and definitely not AC! Enter the power supply. They come in two flavors, linear and switcher.

Linear Power Supplies

AC rules! It is everywhere. It might seem like the world runs on batteries these days, but AC still has the majority foothold. Back when Edison and Tesla argued over what type of electrical power distribution we should have, I'll bet they had no idea of the type of integration that would occur in the world of electricity over the next 100 years.[19]

One thing they did know about was the transformer. The basis of the transformer is AC current. Put AC into one side of the transformer and, depending on the ratio of windings, you get AC out the other side. So, put 120V AC into a 10-to-1 ratio transformer and you will get 12V AC out (minus heat losses due to the resistance of the windings).

The basic transformer is a very simple design. It is coils of wire on hunks of metal. That makes it robust. A transformer is a perfectly acceptable way to change the voltage of an AC signal. Transformers are used to jack the voltage way up to minimize losses over long wires, and then they are used again to bring the voltage back down to something safer to bring into your house.

They further knock the voltage down again in millions of products, but at that point they still output an AC signal. However, most of our chips want a DC signal, so what happens next? It goes through a rectifier. There are two commonly used options: a *bridge rectifier*, shown in Figure 4.23, and a *center tap rectifier*, shown in Figure 4.24 on the next page. Note how this uses two fewer diodes and another wire to the transformer, yet the rectified output is the same. Notice the "bumpy" DC output?

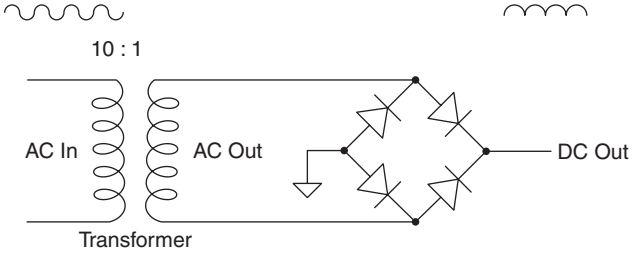


FIGURE 4.23
Bridge rectifier.

[19]Man, did their argument get heated! To the point of electrocuting cats, that is. I won't get into details but point the reader to a great biography of Tesla, called *Man Out of Time*.

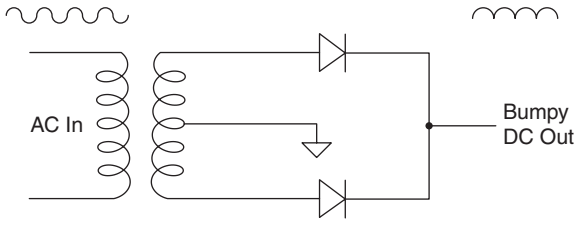


FIGURE 4.24
Center tap bridge rectifier.

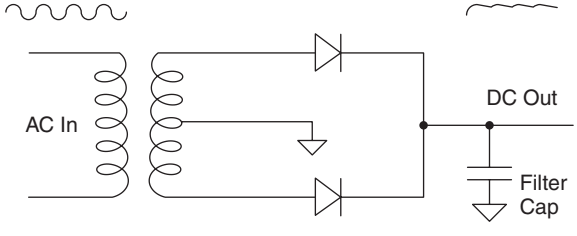


FIGURE 4.25
Center tap bridge rectifier with cap filter.

The output at this point of either rectifier is still too "bumpy" to be of much use to our sensitive DC circuits. The next step is to add a large filter capacitor to smooth out the bumps, as shown in Figure 4.25.

It's time to learn the principle of output impedance. Every power supply has it. The more current you pull out of the circuit, the bigger issue the output impedance is. Remember that Ohm's Law says that as current increases through an impedance, the voltage drop across it increases. This means that the voltage at the output will drop as load increases. To further complicate things, the rectifier in this circuit will contribute to an increased ripple voltage on the output as load increases.

So, two important things affect the voltage on the output of this circuit: the voltage going into it (which on most AC circuits can vary 10% or more) and the amount of current being drawn, increasing voltage drop and voltage ripple.

This is important to know as we feed this into the next part of the circuit, called a *regulator*. The regulator is a part that adjusts its output to maintain a constant voltage in the face of a changing load and a changing input voltage.

A linear regulator typically has a voltage reference (like a zener[20] diode) inside it running on a small current that isn't disrupted by the load. It uses

[20]Zener, zener, zener, man that is a fun word to say! Way more fun to say than a word like coulomb or Schottky.

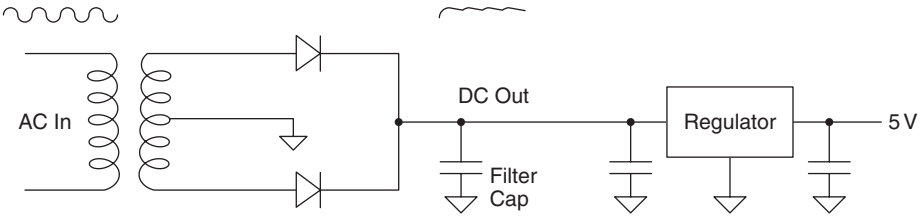


FIGURE 4.26
Typical linear regulated power supply.

this reference and a negative feedback loop to control a transistor or other part inside to maintain a constant voltage at the output. This gets you to the nice DC voltage that your IC wants. The whole circuit from the wall looks something like Figure 4.26.

There are a couple of important things to know about linear regulators. They have a minimum input voltage. If the input voltage falls below this rating due to circumstances described before, the output will fall out of regulation. If this happens you can get ripple on the power supply to your chip. If it is small enough, you might never notice it, but if you have some high-gain circuits picking up AC noise, check out the power supply for problems first.

The other often-overlooked important spec is the power rating of the regulator. A regulator can only handle so much power, even with a heat sink. The power being dissipated by the regulator is the current times the voltage drop across the regulator, *not the voltage at the output!* There are many other specs you should review in the datasheet, but these are the most important and often overlooked. Check them first. You can use linear regulators in any DC-in, DC-out situation. They will do very well in most cases and, to top that, they are very simple and robust circuits. Use them whenever you can. There is nothing wrong with this technology in certain applications, but if you need more efficiency or maybe less heat, you should consider a switcher.

Switchers

A type of regulator and power supply rapidly gaining footholds over the older, linear designs is called a *switcher*. As implied by the name, the switcher regulates power to a load by switching current (or voltage) on and off. In this book we will focus on the current method. (Don't forget, however, that current and voltage are invariably linked, as Ohm proved so well.) The secret to these supplies is the inductor, and the secret to understanding an inductor for me is to think in terms of current. In the same way a capacitor wants to keep voltage

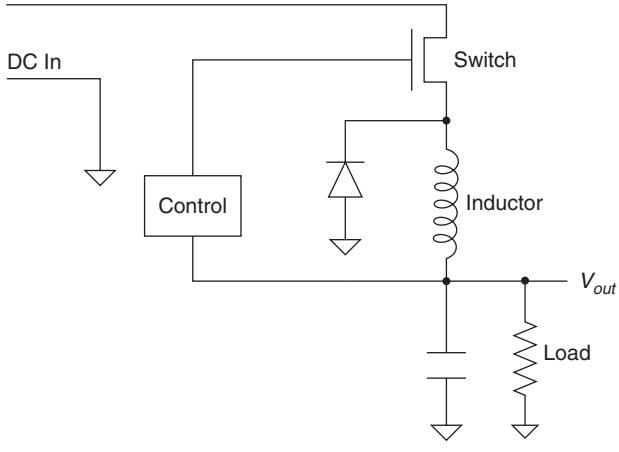


FIGURE 4.27
Basic switching buck converter.

across it constant, an inductor wants to keep the current flowing through it constant as well.

DC IS WHAT WE START WITH

Switching power supplies are DC-to-DC converters. Even those that have an AC input create a DC bus, using a rectifier circuit before implementing a switcher. You will see switchers replacing just the regulator in our earlier circuit working off a DC bus voltage that has already been stepped down by a transformer. You will also see switchers that use rectified voltage right off the AC line and drop and regulate all in one step from 120 V down to 5 V.

The most basic current-switching supply I know of is the buck converter. A buck converter will knock a DC voltage from a higher level to a lower level. Figure 4.27 shows the heart of a buck circuit.

First, let's identify the four main parts: the inductor, the switch, the diode, and the load.

FLIP THE SWITCH

Let's start with the load and work our way backward. To begin with, switching supplies like to have a load. Without a load funny things can happen, but more on that later. What the load wants (in most cases) is a constant voltage. If I remember Ohm's Law correctly, one can control the voltage across a resistor (i.e., load) by controlling the current through it, so let's consider the flow of current in this circuit. We will begin with the switch closed. With the switch

closed, current will flow through the inductor into the load. The current will rise based on the time constant of the inductor and the impedance of the load. Since the current rises, so does the voltage across the load. Assume now that we have a circuit that is monitoring the voltage across the load, and as soon as it gets too high it opens the switch.

Now what happens? First, remember this fact. Just as a capacitor resists a change in voltage, an inductor resists a change in current. When the switch opens, the inductor tries to keep the current flowing. If there is nowhere for it to go, you will see a large voltage develop across the inductor as the magnetic field collapses. In fact, at time = 0 the value of this voltage is infinite or undefined, whichever suits you. That doesn't happen in this case due to the diode and the load.

The current flows into the load, and the reason it does so is because of the diode. Consider it this way: Current wants to keep flowing out of the inductor and into the other side of the inductor. Without the diode there would not be a path for this current to follow. However, with the diode, this current is pushed through the load. So now the switch is open, and current is still flowing into the load. This current starts out at the same level it was when the switch opened (an inductor wants to keep current the same, remember!) and it decays from there. As the current falls, so does the voltage. Of course we still have a circuit monitoring the voltage across the load, and as soon as it gets too low, it closes the switch again.

Voilà, the process starts all over. There are two important things I noticed once the pieces fell into place in my head. The first is that this control circuit I just described can be implemented with a simple comparator and a little hysteresis. Of course, that would lead to the frequency of the switcher being determined by the value of the inductor and the impedance of the load. That might or might not be a desirable trait. The other thing I realized was that when you first turned it on, the circuit would want to slam the switch shut and keep it there for a long time while current builds up in the circuit. Are you beginning to see why switchers need a load?

Luckily, others much smarter than I have dealt with these problems already. That is why you hear terms like *soft start* and *built-in PWM* when you start studying switching supplies.

Some Final Thoughts

Since designing switching supplies, getting them stable, and dealing with the inductor specs can be a bit demanding, technical, and tedious, all sorts of industry help has sprung up in the effort of various companies trying to get you

to use their parts. You will find design guides and even Web design platforms out there to help you build a switcher for your design, and I highly suggest you take advantage of them.

These days you will often find all the brains, switching components, and feedback parts in one cute little package,[21] making the design nicely compact and small. You can make switchers that boost voltage as well as the buck versions, and some that even go both ways, but ultimately they rely on the fact that the inductor wants to keep current flow the same. We will save the more in-depth review for another book on another day.

The best thing about switching supplies is the fact you can get by with relatively little copper and attain very high efficiency (meaning less heat). The reason for this is that the decay rate of the current in the inductor depends on the size of it, but if you switch it faster, the average current and thus the average voltage is still maintained. So you can get by with much less copper, especially for larger current draws at low voltages. The efficiency is good because much less power is spent heating the copper in the small inductor than in an equivalent transformer design. However, all this comes at a price. Switchers are known for their high-frequency noise that has disrupted many a sensitive analog design. But who cares about analog anymore, right?

Thumb Rules

- Make sure the lowest dip on the ripple voltage doesn't go below the minimum input of the regulator.
- Check your supply at ±15% of the AC input signal.
- Linear regulators dissipate heat/power based on the current times the voltage from input to output (i.e., across it).
- Switchers exploit the fact that inductors want to keep current flowing even when the switch is open.
- Switchers are more efficient and create less heat but generally are more finicky to set up.
- Linear supplies are very quiet when it comes to EMI.
- Switchers tend to be very noisy when it comes to EMI.
- Switchers need a minimum load to work correctly.

[21] I know, only real nerdy engineers would think an IC could have cute packaging, but I have never denied my nerdhood.

WHEN PARTS AREN'T PERFECT

Before we get into the problems that parts can have, we need to introduce the concept of an equivalent circuit. It is pretty simple: to create an equivalent circuit, you represent all its idiosyncrasies with combinations of perfect components. This is good for two reasons: First and most obvious, it makes it possible to model the effects of the imperfections. Second, and most important in the World of Darren, is that seeing the combinations of the parts that make up a real component makes it easier for you to apply the basic understanding of the perfect parts to grasp what the real part is doing.

Everything Is Everywhere

The basic three electrical components are like sand at the beach. They get into everything. In a way they are more prolific than sand in your sandals since the effect of one basic component can be found in another. This fact is one of the most common causes of error you will have between the way the equation predicts a circuit will work and the way it actually operates. Chalk this up as one of the reasons datasheets are so important, even the ones that describe the most basic components. Datasheets will characterize the components, describing these error sources.

Most texts call these effects *error sources* since they are what makes the difference between a perfect or ideal component and what you actually have to work with. There are other types of error sources in every component, and we will discuss a few of them later on, but those pesky R, L, and C in some combination or another are pretty much everywhere. (I hope the reason for drilling the basics of these components is becoming more and more clear. It is appropriate to experience an "Aha!" moment right now and say to yourself, "Now I see why I need to know those basic parts by heart!")

The most general guideline to follow when you are looking for error sources is to ask yourself the following: Is this error source enough to account for the effect I am seeing?" Let's consider a diode as an example for a moment. A diode has a bit of capacitance when it is reverse-biased, typically in the picofarad range.

Consider the circuit shown in Figure 4.28 on the next page. If you hook your scope lead to the output, and flip the switch, you see what's shown in Figure 4.29. Since there is capacitance in this diode, an RC curve is what you would expect to see in a situation like this, but is it really due to the error source in this diode or is it caused by something else?

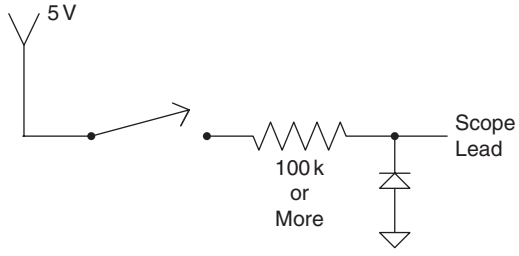


FIGURE 4.28
Circuit to examine.

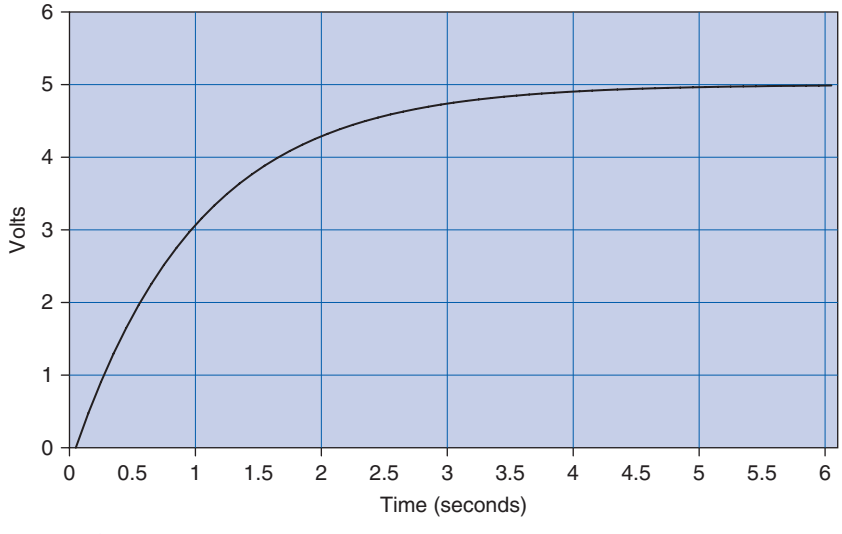


FIGURE 4.29
RC curve seen on your scope.

Here is where the datasheet comes in; looking at the specs on the diode you are using, you find out that this capacitance is typically 100 pf. Plug that into the equation for the response time of an RC circuit:

$$\tau = RC$$

The number that pops out is 10 μS. Taking a look at the scope again, you now pay attention to the time it takes for this curve to get to about 63%, remembering that is about how far this curve gets in one time constant, or tau. Being such an astute scope operator, you use the cursors on your tool and you measure a time of about 1 second for the signal to get a little over halfway to its final value. "That doesn't make sense," you should be saying to yourself, "If the diode is responsible, it would have to be much faster."

The moral of the story is to expect every component to have some amount of the basic three, but understand the magnitude so you can decide if it is causing the effect you are seeing.

Error Sources, Ideal vs. Real

In any circuit you design, there will be sources of error—things that simply aren't perfect, sensors that are off a little, parts that aren't quite to spec, or any myriad of problems.

What do you do about it? Nothing, if the error source isn't causing you any problems. For example, a leaky cap might not really matter if you have plenty of power available. However, if the circuit is running on batteries you could have a problem on your hands. First and most important, determine whether the source of error is an issue or not before you go about trying to solve it.

Once you figure you have a problem on your hands with a particular source of error, there are three ways to deal with it:

1. *Get a better component.* It's tough to plug the hole in a leaky cap;[22] it isn't like the boy at the dike—you can't put your finger in the hole and stop the leak. Sometimes your only choice is another component. In this case you might specify a tantalum cap instead of an electrolytic. Consider, however, that often the better component costs more, so spend wisely, not indiscriminately. Do note, however, that this is usually the quickest way to solve the problem since the design doesn't have to change.
2. *Shore up the weak component with another component.* For example, the frequency response problems with electrolytic caps can be dealt with by adding another cap in parallel, a smaller one that has no problems with higher frequencies. (You might have noticed regulator reference designs do just that to assure a stable output. Now you know why.)
3. *Design the error out.* This approach will take the most engineering effort, since the goal is to change the design so that the error is no longer significant. The proverbial op-amp is an example of this type of effort.[23]

---

[22] A very common source of error in a capacitor is a DC current flow. Remember, the ideal cap will block all DC signals. You can think of it as a large resistor in parallel with a perfect cap. It is common enough to have acquired its own slang term: If this current flow is significant, the cap is said to be *leaky*, because DC current seems to leak through it.

[23] The whole point of the op-amp was to eliminate error sources in designing transistor amplifiers. It was a pretty cool idea, but it did take some real work!

Now that we know how to deal with the problem, let's look at some common parts and typical sources of error. This will be an overview based on personal experience. It is no substitute for curing insomnia with a good datasheet.

Resistors

I would have to say that resistors are the most stable and predictable of the three basic components. Carbon film resistors have very little inductance or capacitance. It is rare that you will have a problem with this unless you are dealing with radio frequencies and high clock speeds. In most cases the effect of the PCB design will be worse than the resistor itself.

The biggest issue with these common resistors will likely be heat. Exceeding or coming close to the wattage rating of these parts will make them vary significantly from their nominal value, so it is a good idea to give yourself plenty of headroom with these resistors.

Another common resistor typically used in higher-power applications is a wire-wound coil with a ceramic block molded around it. In this case inductance can be a significant effect since there is a coil of wire and, as we know, coils of wire make inductors. There is a whole industry of low-inductance power resistors that you can get to work around this problem.

Capacitors

In my personal experience, I have never seen a cap that even comes close to being perfect. A perfect cap would not heat up, but in fact they do. The natural conclusion you should come to is that capacitors have some type of resistive component. In fact they do, and it is called ESR, or *equivalent series resistance*.

According to the equations, a 10-μf cap should have nearly the same impedance at 100 K Hz as a 0.1 μf cap does, but alas this is not the case. That is why you often see a large cap with a small cap next to it on a power-supply circuit. Nearly all caps vary capacitance over a frequency range.

Big electrolytic caps often "leak" like a sieve. There is no particularly easy way to deal with that. You have to live with it or get a better part. Believe me, if you are trying to make a really low-power circuit, the last thing you want is a cap spilling electrons all over the place.

One other thing I had to learn the hard way is that a cap only meets the rated capacitance when at the rated voltage. Sometimes overrating the voltage on the cap too much can leave you with a different capacitance than you expect.

Polarized capacitors will act like a diode if you don't bias them according to their markings. Many caps will vary 20% over their temperature range; you might not want them next to a power resistor on your PCB.

The moral of the cap story: You need to peruse capacitor datasheets carefully when you are picking them for a particular application.

Inductors

Since these are most often coils of wire, you might imagine that resistance is one of the most common sources of error in an inductor, and you'd be right. Resistance is a major source of error in inductors. This usually causes heat and power usage that you might or may not want. Minimizing the current flow through the inductor makes the resistance less of an effect and is something you might be able to do at the design stage.

Many inductors are warped around some type of ferrous core. An effect called *core saturation* occurs when the magnetic field exceeds the amount the core can handle.

There are some capacitive effects between the coils of wire, but they are so small that we will ignore them in this text. If you are cranking out the gigahertz needed to make this important, you are probably reading a book about this stuff written by someone much smarter than I am.

Semiconductors

One of the things that every diode, and every semiconductor based on the diode, has in it is a voltage drop. For example, if your transistor amplifier doesn't see a base voltage over 0.7 V, you won't get it to work.

Rail-to-rail op-amps are more expensive than their predecessors because they employ circuitry that eliminates these voltage drops so that outputs and inputs can get to their power rails.

In the datasheet of these parts, you should look for output impedances and capacitive effects. Inductive effects are generally small and insignificant in semiconductors.

Heat can also cause error in semiconductors. It generally affects the internal resistance and can cause avalanche[24] failures. It also seems to me that the most

[24]Like an avalanche, when it starts to fail all hell breaks loose, usually resulting in an interesting smell.

often overlooked part of the design is heat dissipation. The same engineers that can easily calculate the wattage needed for that specific resistor value will overlook the amount of heat dissipation in a semiconductor. Take the current though the part times the voltage drop across it and you will see how much power is being dissipated.

The world of semiconductors is so widely varied that there is no way this overview can be anywhere near comprehensive. I have to sound like a broken record (or should I say scratched CD?) and tell you to refer to the datasheet.

Voltage Sources

What would cause a voltage source not to maintain the voltage output? Let me give you a hint: When put under load, a voltage source will heat up. So what creates heat? You got it: resistance. A voltage source has an internal resistance. A battery is a good example—see Figure 4.30.

As current is applied to the load, the voltage drop occurs across this internal resistance, just like the voltage divider rule says it will. This resistor inside heats up just like one on the outside does, making the voltage source warm. If the source doesn't compensate for it, you will see less voltage at the output.

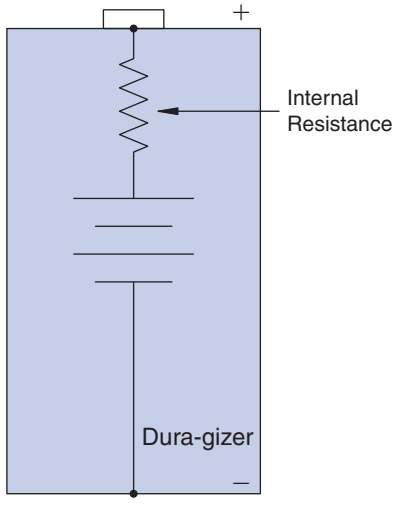


FIGURE 4.30
The Dura-gizer; now that is one tough battery!

When Parts Aren't Perfect Conclusion

Parts simply aren't perfect. I have seen motor bearings wear out prematurely due to capacitive effects and have seen caps overheat and pop their tops. Truly the

best thing to do is keep looking at the datasheet. Parts engineers do their best to characterize the deficiencies of the part and put it in the datasheet for you.

Thumb Rules

- Always ask, is the error source in this component enough to cause the effect I am seeing?
- If the source of error isn't large enough to be an issue, forget about it and move on.
- When fixing errors, get a better part, shore it up, or design it out.
- Caps vary with frequency.
- Inductors have internal resistance.
- Semiconductors have voltage drops and heat issues.
- Voltage sources have internal impedance.
- You can't study the datasheet too much.

ROBUST DESIGN

Most engineers want to over-design, give themselves plenty of headroom, and use parts that are double or triple the spec they need. Usually the manager is there, saying, "It needs to cost less or it won't sell" or "Do you really, really need that part?" To be honest, the answer lies somewhere between these extremes.

Can You Tolerate It?

Let's start with a completely general off-the-wall statement that you might hear from someone with a sharpened, somewhat devilish hairdo. "A robust design handles two things well: the inside world and the outside world." A look of consternation scrambles across your forehead. "What in the world does that mean?" you ask yourself. Let me proceed to enlighten you on this bit of pointy-speak.

The inside world is all the parts that make up the design. In any production process, these parts will vary in specification. The question to ask is, will the circuit operate correctly over the tolerance ranges of the parts? If the answer is yes, the design is robust internally. The inside world is good to go. Don't assume, however, that only electronic components have tolerances. This point is best taught by example. In a design I worked on some time ago, we were using an

optic sensor to detect the motion of a belt. We correctly analyzed the tolerance of the opto, but as we began testing on a pilot run we discovered that the belt we were using varied in opacity. If the optic sensor was at the high end of its tolerance and the belt was at its most transparent, the signal from the sensor wouldn't get high enough to guarantee that the logic input would read it correctly.

In a production run, a problem like this would appear as a random failure. This type of failure is known as a *tolerance stack-up*. It occurs when the stack-up (the additive effect of the variations) of two or more components combine to create a failure. It is more difficult to analyze than a single component tolerance issue. Probably the best way to preempt this type of failure is with the help of simulators. Take caution, though: Make sure that your simulation accurately represents the design with nominal perfect components before you start running tolerance analysis on it. (See the section on simulators for more suggestions.) The great thing about a simulator, though, is the ability to vary all the components over their tolerances and see the effects without building a whole bunch of parts. You can then adjust your design and component specs to increase the robustness of the product as far as the inside world is concerned.

Now the outside world is a different animal. A good design can handle the things the outside world throws at it. In the electronic realm all sorts of interference can disrupt your design. I once read an article that described something called a rusty file test. After the engineer was done with the part, he would plug it into the wall and plug in a home-built test fixture next to it. It consisted of a wire from neutral connected to a file. The hot wire had a bare end that he would proceed to rub up and down the rusty old file, sparks flying everywhere.[25] If the circuit passed this test without a hitch, he figured it was good to go. This is known as EMI, or *electromagnetic interference*. It really is a whole topic unto itself, so I have dedicated a chapter to it. Skip ahead to it if you can't handle the suspense!

Don't limit your focus on the outside world to electrical interference. There are many cases where other things can cause a problem. Vibration, for example, can cause traces on a PCB to crack and solder joints to become faulty. Increased humidity can swell a cheap PCB, causing mechanical deformation and cracked connections. It can also combine with debris to create electrical shorts on circuits that you don't want shorted. Temperature can be particularly tough on electrical

[25] It was written by Ron Mancini in EDN, but I have to say: Do not try this test at home. There are much safer ways than the procedure described; I mention this test because it creates a vivid picture of the junk out there that is trying to mess up your circuit.

components. You should review the temperature range your circuit will be subject to and compare that to the specs in the datasheet. Don't forget to include the operating temps of the device you are using in this analysis. For example, power components usually get pretty warm just operating. Toss them into a 70-degree ambient and you could easily push them over the max temperature spec.

How do you go about making your design robust externally? There are several approaches to take:

- The most important, in my opinion, is doing everything you can in the fundamental design to get it to handle the environment it is in. Often a few changes to the PCB layout itself can make a circuit handle EMI better than putting all the shielding around you can fit. Larger traces can combat mechanical deformation, and a few well-placed holes can help manage temperatures.
- Reading, reading, and rereading the datasheet for the component you are using is probably the next most important thing you can do. The more you know about the parts you are using, the better you will recognize things that might upset your design.
- The third and most extensive effort that will help you is to test, check, test, and retest the design. You need to recreate the environments that it will be subject to and see what happens.

Now, to top it off, you can have a situation where the problem is a combination of the tolerance of the internal design and the environmental effects it is subject to. These situations are nearly impossible to predict and are often simply discovered in the course of business. There is only one thing you can do about that: Figure out what is needed to prevent it, make the change, and document it for future use on similar designs.

I recommend that every engineer and engineering group keep a document of design guidelines[26] where you write down those rules of thumb that you discover along the way. Don't just write it down, but read it regularly to keep those things you have learned fresh as you do each new design. This alone can be a powerful tool. Some years back I took over an engineering group. When I first started managing it, it seemed like we were always being called to the production line for some weird problem or another. We spent more time chasing problems than engineering new products.

[26] I like to call them *gauntlets*. If the design can run the gauntlet of passing guidelines and tests, that is when I deem it good enough.

We began a focus on robust design principles, and one of the first things I implemented was the design guideline documents. Every time we found a new design rule to follow, we wrote it down and referred to it regularly so that it would be implemented with each new design.

Over about a three-year period, those urgent calls to production began to drop off. We went from spending over 50% of our time in production support to spending less than 10%. A couple years after that, we were spending less than 1% of our time dealing with production problems. Considering that we were moving tens of thousands of products out per day, it was a great achievement. Months would go by without a call, where before we got calls every day. When problems did occur you could nearly always trace back to a guideline that we had written down and simply neglected to follow. The hard part became referring back to those documents each time we created a new design. That being the case, I suggest you try not to let your guidelines get too large. The bigger these documents, the less likely you are to read through them. So try to keep them to a few pages, since they will have a tendency to grow a lot.

In an effort to quantify what the outside world can do, many standards have been written. They are some great yawners (meaning they will knock you out in about 5 minutes of reading); however, they can give you some real insight into what your design will be subject to from the outside in. I'm referring to documents like IEEE 62.41, which describes the world of EMI, or UL 991, which describes how to make a control safe. The list goes on and on. Do a little research into what you are working on and see if someone has written something about it. If your boss doesn't understand the need for time to do this, show him this paragraph:

> Boss, it might seem like nothing is getting done when the engineer is sitting there reading, but trust me, this effort can save you millions in production downtime, so give your engineer a chance to succeed and you will not regret it. Engineer, this doesn't mean that you should just read and never design anything; I would limit this research to about a 10 to 20% ratio of design vs. research; double it if you are doing something you have never done before.

Reading these documents works particularly well if you are tossing and turning all night as you try to figure out what is wrong with your design. I would keep them by the side of my bed. That way I could learn some more for a few minutes and also get some sleep. They not only help with the design, they are a great cure for insomnia!

Learn to Adapt

Have you ever finished a product design after which some change was required that you desperately wished you had been told about at the beginning? Have you ever said, "If you had just told me sooner, this feature would have cost half as much to add on now"? You might even have had your boss say, "Why didn't you do what I told you to do?" (when you did exactly what he or she wanted). I'll let you in on a secret: Most of your pointy-haired bosses don't want you to fail. They just want to ship a killer product so their bonus will be bigger.

They don't tell you sooner because they don't know sooner. They try to guess what the customers want and give it to them. In his mind your boss didn't say, "Do such and such a product; he said, "Make this product successful." As companies chase the market around, new products are developed, changed, and changed again. I call this Management Always chasing the Market Around, or MAMA for short. (There is nothing like an acronym if you want a point to be remembered. I predict that some day in the far future acronyms will be the prime method of communication due to their efficiency!) Now, because of MAMA, many engineers experience consternation when their product definition changes.

In the world of consumer products, this is bound to happen often. When I took over the engineering group in the first company I worked for, this particular frustration was often felt. As I worked with the various designers in the company, I found that it was possible to anticipate these changes and prepare for them. When you get good at this, you can respond to changes easily and quickly, and you can also develop a number of derivative products quickly and inexpensively.

Modular Design

One of the most important things that you need to do to anticipate change is to modularize your designs. Here, hardware designers can take note from their counterparts in software design. Good software engineers build blocks of code that can be used and reused again and again. However, I often see hardware designers start with a clean sheet for every new design.

To make your modular design work for you, you must evaluate the products you are designing. Are there any components that are commonly removed and installed on various designs? What parts are common to all or most of your products? Sit down and draw lots of block diagrams and ask yourself, "Is this a part that needs to be easy to take on and off?" If it is, it might be a candidate

for a separate PCB or a section of the PCB all to itself. In a line of stereo products, for example, you keep the tuner section separate from the pre-amp and so on. (On a side note, this often improves the robustness of a design as well.)

A great advantage of this modular approach is the way it can accelerate the development process by using separate engineers on the various modules. It also allows you to upgrade or improve parts of the design without redoing the whole thing. Most important in my world and best of all, it makes it easy to change a feature when your boss decides he really didn't want that there on this particular model.

One word of warning, however: You need to be careful what parts you choose to modularize. Too many modules can add up to extra cost in every product you ship. *Make sure your choices make sense.*

Anticipate Changes

Try to get involved in the creation process so that you can see various phases of evolution the product design has gone through. Often, changes that are made will be back or forth on this evolutionary path. Keep asking yourself, "Where else could this be used? How would I change it to work there?"

Look for places where a part seems to be missing. For example, say that you are asked to make a PCB with a row of LEDs that look like Figure 4.31. Say "Great, no problem," and then create a PCB with this row of LEDs, as shown in Figure 4.32, and simply do not install the missing one for now.

Don't hesitate to tell your coworkers or boss what you are doing. They can be a great asset in anticipating changes they might come up with later. The bottom line is that it takes a tremendous amount of work to redesign every product every time, but if you can develop effective modules and anticipate changes when you are engineering the product, you can bring things to market faster

FIGURE 4.31
Row of LEDs management wants.

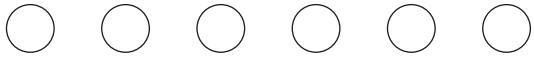


FIGURE 4.32
Row of LEDs you actually put in.

and cheaper than anyone else. A nice benefit of this type of anticipative design is that when you are asked to develop a similar product, you have all the pieces in place. You simply add or subtract the required feature and are done with it. Finally, best of all, MAMA will not drive you berserk!

One Last Word of Caution

It is possible to go too far with this philosophy. Don't try to make your design so universal that it comes at the expense of getting the product to market or adds so much cost for all the options that it is no longer viable. Remember, there is also a chance you will never use the option you built in, so choose wisely, young Jedi.[27]

Thumb Rules

- Read the datasheet.
- Consider tolerances.
- Know the environment.
- Test, check, and retest.
- Make your own list of Thumb Rules or design guidelines.
- Do research on standards or guidelines that exist for your product.
- MAMA can be frustrating.
- Modularize the design.
- Anticipate changes.
- Don't go too far.

SOME OF MY FAVORITE CIRCUITS

Every engineer has their favorite batch of circuits, and I'm no exception. There are tons of circuit cookbooks out there that show how to implement no end of cool features. There are so many that you could spend all your time searching them and never getting anything done. I suggest you develop your own favorite basic circuits that you know well and intuitively understand. This is simply an

[27] Do Jedi mind tricks work in the cooperate world? I think so. Now, that is a cool idea for a book. Email and let me know if you would buy it. If I get enough responses I definitely will explore that idea!

extension of the Lego philosophy that we discussed way back at the beginning of the book. Here are a few of my favorites. These are in addition to all the circuits I have used as examples up to this point. One reason they make such good examples is that they are so useful.

Hybrid Darlington Pair

Cool application note: using two transistors to switch a signal level *Vcc* PNP switched by NPN.

Figure 4.33 shows a handy circuit that switches a higher-level voltage with a lower-level one. Say, for example, you have a micro with a 5V output and you need to drive a 12V load. For a reason you can't change, you have to switch the *Vcc* leg. In this circuit you turn on one transistor with a 5V signal, which in turn activates the other transistor, switching the higher voltage to the load.

This works because the transistors are current driven; when you shut off the current flow to the PNP transistor, it shuts off regardless of the voltage. Another plus is that this circuit has Darlington-like properties without one of the downsides. You won't need a lot of current to the input to switch the output and, unlike a traditional Darlington pair, the voltage drop across the output is much smaller. You don't have two series base junctions to contend with at the output. If you still don't follow, try a little ISA[28] on it.

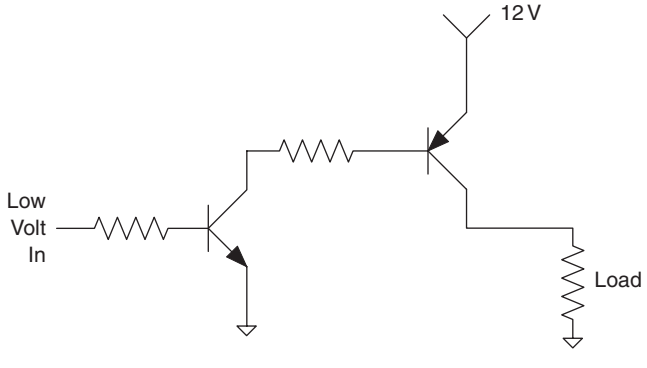


FIGURE 4.33
Vcc PNP switched by NPN.

[28] Intuitive signal analysis (ISA). I still hope to someday cement my legacy in an acronym.

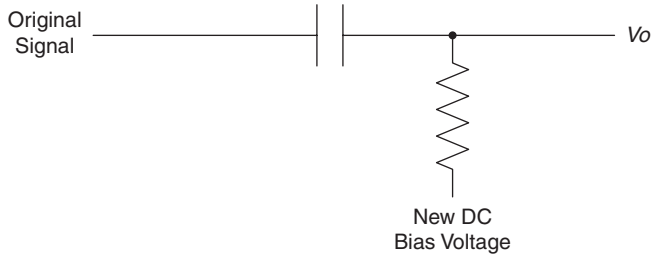


FIGURE 4.34
Change the DC bias on an AC signal.

DC Level Shifter

This is really the high-pass filter that we have already studied but with a slight twist, as shown in Figure 4.34. Instead of ground, we hook the resistor to a reference voltage. Since DC has a frequency of zero, only the AC component will pass and in the process a DC bias will be applied to the signal. Make sure that you don't size the cap and resistor so that the signal you want is attenuated.

Virtual Ground

Using the voltage divider as a reference, the op-amp becomes a voltage source with the output matching the voltage at the divider—see Figure 4.35. This can

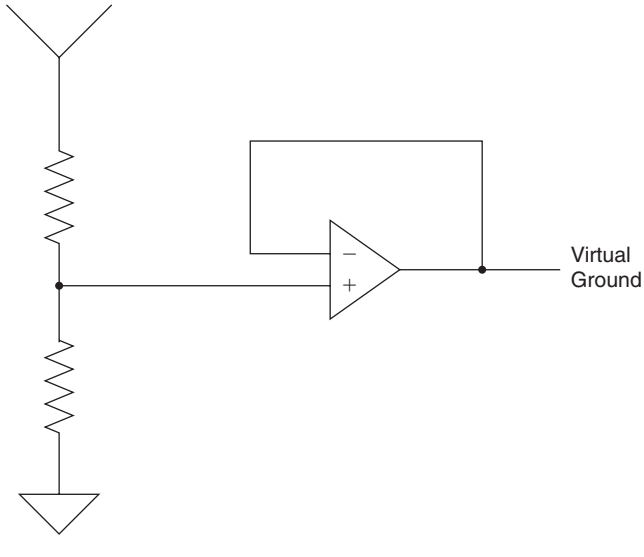


FIGURE 4.35
Create a "ground" at any level you want.

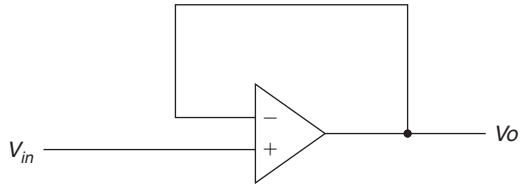


FIGURE 4.36
Voltage follower.

be very useful when you are trying to handle AC signals with only a single-ended supply circuit.

Voltage Follower

As Figure 4.36 shows, this one is mighty useful when you're trying to measure a signal that is easily affected by load. *Vi* = *Vo*, but, best of all, *Vi* isn't loaded at all, thanks to the buffering effect of the op-amp.

AC-Only Amplifier

Figure 4.37 shows another great circuit that works nicely in amplifying AC signals with a single-ended supply. It also has the benefit of not amplifying any DC signal components, keeping things like DC offsets from making your signal rail. This happens because of the cap in the feedback circuit. Since the cap only passes AC current, DC signals see that point as disconnected. When the resistor to ground is disconnected, the op-amp acts like the voltage follower in the previous circuit.

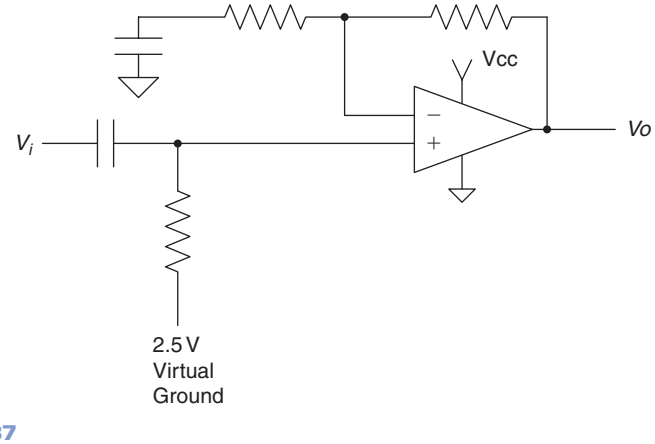


FIGURE 4.37
AC-only amplifier.

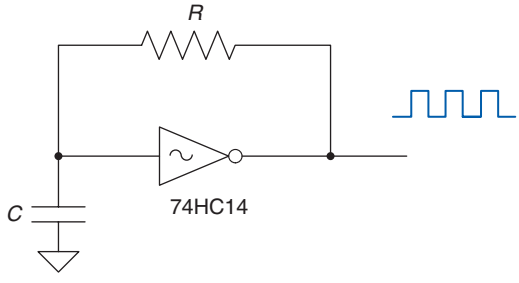


FIGURE 4.38
Schmidt trigger oscillator.

Inverter Oscillator

I saw this in the back of a data book years ago; I think it was a Motorola logic data book. This was way back before the Internet. You used to have to turn actual pages to find this stuff! The way it works is based on the fact that the Schmidt trigger inverter has hysteresis built into the input (Figure 4.38). This makes the output stick at a high or low level till the cap on the input charges to the threshold voltage that trips the inverter. Output flips and everything goes in the other direction, repeating indefinitely. Adding some diodes to the charge and discharge path can affect the duty cycle of the output.

Constant Current Source

Using negative feedback, the op-amp tries to maintain the voltage drop across *R* input. Even if the resistance of the load changes, the drop across *R* input stays the same. According to Ohm's Law, keeping *R* and *V* the same will keep current the same, too—see Figure 4.39 on next page. Remember, though, this current control has operational limits; it can only swing the output voltage so far to compensate for load variance. Once these limits are reached, the current regulation can no longer exist.

GET YOUR OWN—HERE ARE A FEW

I have just a few favorite circuit concepts. Get your own and know them well. You will be better served knowing a few circuit concepts inside-out than knowing thousands superficially.

Following this advice, several readers of the first edition sent in some of their favorite circuits. Without further ado, they are presented next.

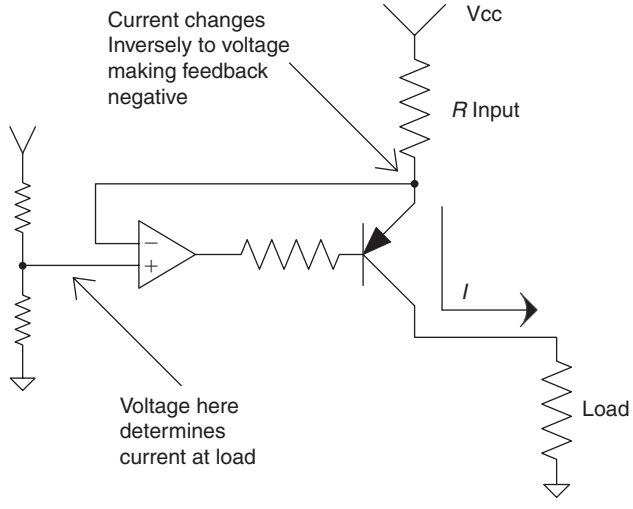


FIGURE 4.39
Voltage-controlled constant current source.

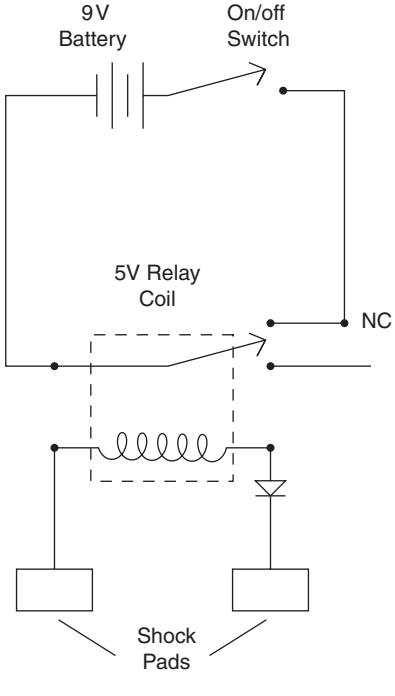


FIGURE 4.40
Toy shocker circuit.

Steve Petersen sent in the circuit shown in Figure 4.40, saying something about being fun for parties and the potential to add a delay circuit to really surprise

someone[29] when they picked up whatever interesting device the circuit was embedded in.

Travis Hayes sent in the diagram of a slick little circuit, as shown in Figure 4.41, that uses the inverter oscillator from my list to drive a voltage doubler circuit. He said it was a pretty slick and inexpensive way to get a higher voltage for an LCD he was using. I'd have to agree!

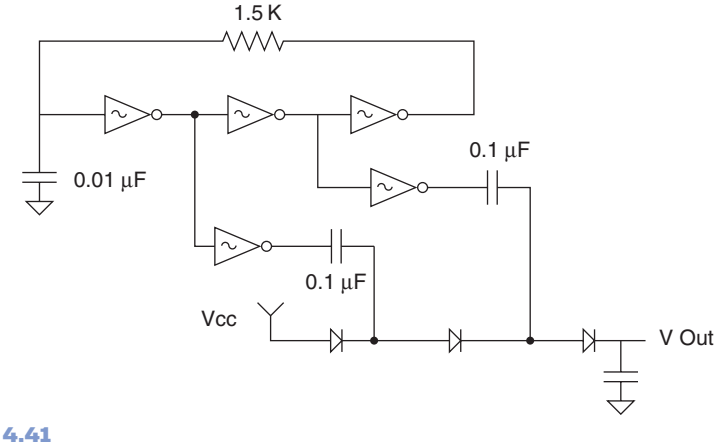


FIGURE 4.41
Inverter-driven voltage doubler.

Alan Tyger just might be as big a fan of op-amps as I am; he sent in the circuit diagram shown in Figure 4.42 on the next page; it uses just such a device to store a piece of information.

Michael Covington[30] sent in the cool circuit shown in Figure 4.43 on the next page; it combines the fun of remote controls with a laser pointer. The 555 acts as a memory cell (not unlike Alan's circuit), but this one has the added bonus that you use a laser to control it. How cool is that! I don't know any engineer who doesn't like lasers, and I'm pretty sure they all have to control the remote when they are home watching TV.

---

[29]I hereby claim no responsibility whatsoever for anyone out there hurting themselves using a design they found in this book when I took the effort in this footnote to say Don't try this at home. We book writers are professionals and know how to do a practical joke without really hurting anyone, at least not too badly!

[30]This circuit was published in the "Q&A" column of *Electronics Now* some time in the late 1990s when I was writing that column for the magazine. The publisher has given permission to republish it elsewhere.

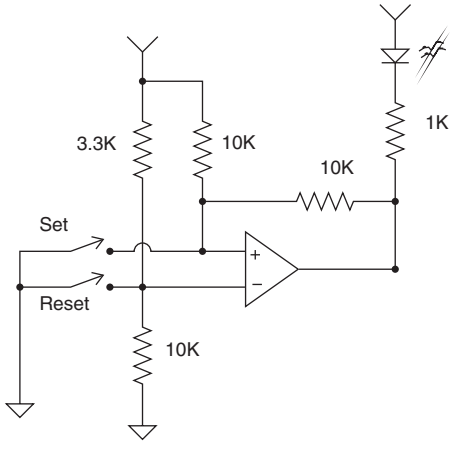
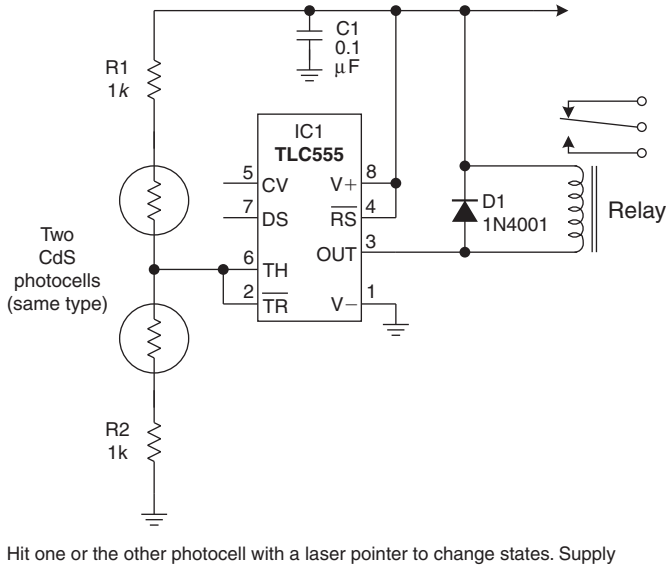


FIGURE 4.42
Flip-flop memory op-amp.



Hit one or the other photocell with a laser pointer to change states. Supply voltage is not critical (5 to 15 V depending on requirements of relay).

Note: Either R1 or R2 (not both) can be omitted to make up for imbalance between the photocells and provide better performance in bright-light conditions.

FIGURE 4.43
Laser light switch.

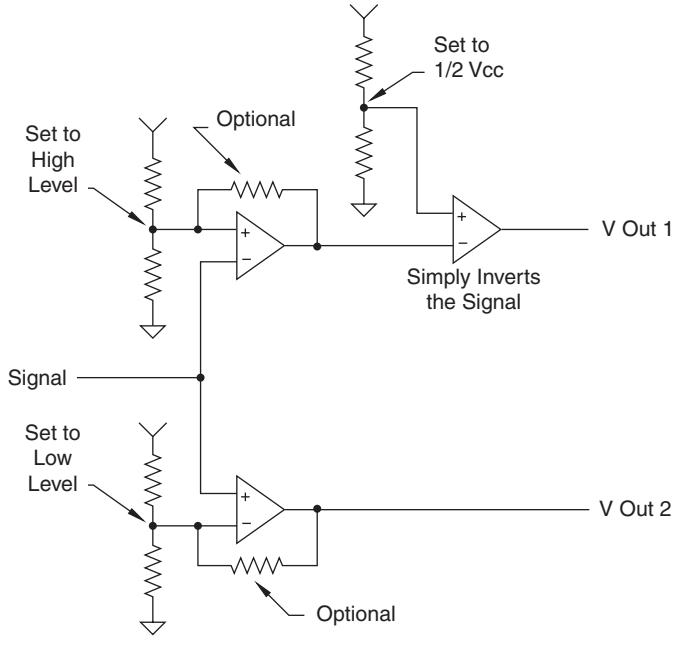


FIGURE 4.44
High-impedance window comparator.

Mike Angeli sent in this cool circuit in Figure 4.44. He said he used it to position a load using a potentiometer feedback (thus the high impedance requirement).

Sam Nay sent in the circuit shown in Figure 4.45, saying he was always fascinated by the ability to transmit data without wires. I'll bet he hooks up the laser-controlled switch that we saw just moment ago. Also, I happen to know

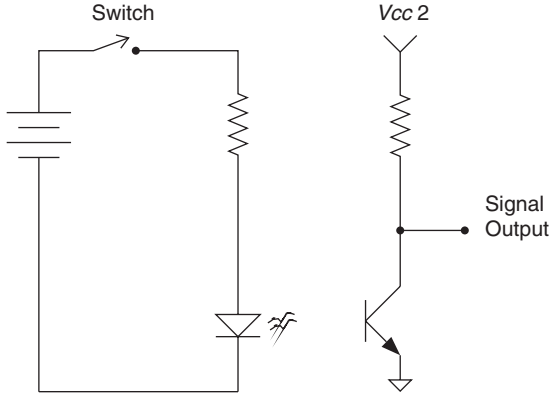


FIGURE 4.45
Optical signal transmission.

of a secret circuit that I am not at liberty to disclose that uses a variation of optical circuits not that different from this one to take biometric readings. Bet you wish I could show you that one, don't you!?

Finally, Mourly Thov sent in the circuit shown in Figure 4.46. He said he just thought it was a slick way to change the DC voltage (and have some power capacity, which could be an issue with the one Travis sent in), so if you find yourself in need of a different voltage that can move some current, try an idea like this one.

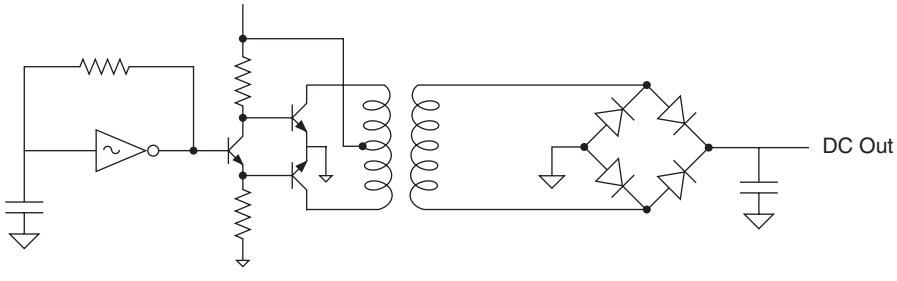


FIGURE 4.46
Isolated DC-DC converter.

On a final note, I have to say that from my communication with these engineers, I think they all fall in the RSP[30] category. Then again, maybe that is just because they emailed me and really liked the first edition of this book. Either way, I thank them for their submissions and completely absolve myself from any responsibility for these circuits actually working. I hope they bring you luck and help you to fill up a notebook with your favorite circuits.

Thumb Rules

- Keep your own cookbook of cool circuits.
- Learn them well.

[30] Look it up in the index at the back of this book. I'll bet you know some RSPs too!

CHAPTER 5

Tools

What tools to use and when to use them? That is the proverbial question. There are innumerable tools available to the electrical engineer. This chapter is not intended to be an exhaustive analysis of various tools but rather a guideline for selecting the proper tool and for setting it up right to get the information you are looking for.

MAKING THE INVISIBLE VISIBLE

One difficulty that electrical engineers have in general is that the electrons they work with are not something you can pick up, touch, or feel. (Okay, I take back the feel part—there are definitely voltage levels that you can feel!) In reality, you infer the electrons' existence based on how they affect objects. You don't see the current flowing in a light bulb; you see the heat generated by the current flowing in a light bulb. Given this, tools that can measure various attributes of electricity are invaluable in electrical engineering.

Meters

Ahh, the lowly meter—probably the tool you will use most often in your quest for electrical engineering excellence. The first rule of thumb in using a meter (and this applies generally to all the tools you use) is to have some idea of what you are looking for. For example, if you are trying to read an AC signal, don't set your meter to DC. This might sound overly simplistic, but I believe that poor tool setup is the most common mistake made by the average engineer.

You can extrapolate from this rule to solve other misreading problems. This leads to a second rule and a case in point. The rule: Don't trust auto setups implicitly. The case in point was a motor voltage we were trying to read; the voltage across the motor was a PWM signal with a peak of 140 V. We were trying to read the average voltage across this motor with a Fluke 87, but the readings didn't make sense (note the application of rule one). We found that when the meter was in auto-range mode, the brain of the meter was confused by the PWM input. Setting the meter manually to the correct range resulted in an accurate and stable reading.

The two most common signals you will examine with a meter are voltage and current. In setting up a meter to read voltage, remember that you are hooking the leads up in parallel with the signal you are going to examine. When reading current, the meter must be hooked up in series in the circuit. Remember that current is a measurement of flow. Nearly all meters require you to hook the leads into different plugs when reading voltage than those that are used when measuring current. This is so the signal can be routed through an internal shunt resistor across which a voltage is measured and scaled to represent current.

Typically, there is a fuse in the meter to protect this shunt from overload. On some meters the shunt is a different value for different ranges of current.

All meters will affect the circuit they are hooked to, whether they are in voltage mode or current mode. The question you should ask is, "How much?" A typical digital multimeter (DMM) has 1 to 10 M of impedance in the voltage-measuring circuit. As you hook the leads up to the circuit, consider that you are adding a resistor to the same points.

Let's look at the example shown in Figure 5.1. We will assume that our meter has a 10-MΩ input impedance and we are measuring the output of a voltage-divider circuit.

Let's calculate the effect this has on the circuit. We will start by calculating the parallel resistance of the meter and the resistor it is hooked to:

> Hmmm, scribble scribble, nibble on the pencil eraser, mumble to myself, I just learned that rule, it's the product over the sums, so that would be (1 K * 10 M)/(1 K + 10 M) or 0.9999 K.

Now we apply the voltage divider rule. More humming, more scribbling, and we see that without the meter the output will be at 2.5 V, but with the meter

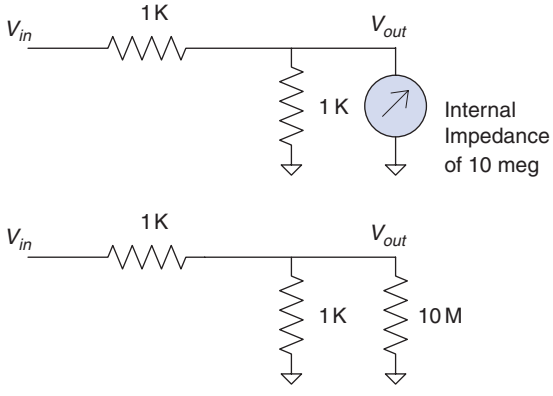


FIGURE 5.1
Equivalent circuit of a meter on a 1-K voltage divider.

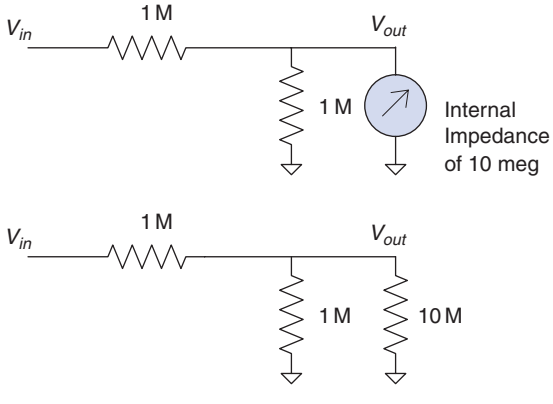


FIGURE 5.2
Same meter on a 1-M voltage divider.

the output will be 2.4999 V. So we will probably all agree that the meter does not have a significant effect in this case.

Let's change the value of the resistors and see what happens. We will make them 1 MΩ resistors, as shown in Figure 5.2.

The first thing you should notice is that without the meter the voltage output will be the same as the previous circuit. But what happens when you hook up the meter? 1M//10 M (the // marks mean *in parallel with*) gives a value of 909.09 K. Run that through the voltage divider rule and you get 2.3809 V as the output. Do you see how the meter can make a difference? Hopefully, what your intuition is telling you is that the effect of the meter depends on the ratio of

the meter impedance to the impedance of the circuit you are reading. Now try an experiment. Change either resistor in the 1M divider to 1 K and run through the same analysis. You will see that the meter no longer has a significant effect. This is because the overall impedance of the circuit is about 1 K. Thevenin taught us that. If you don't quite follow, now is a good time to flip back to Chapter 2 and bone up on Thevenizing. Make sure you consider the overall impedance of the circuit you are measuring when you're determining the effect a meter will have on your circuit.

Scopes

The primary two controls on a scope are just like in the old TV show, *Outer Limits*: "We control the vertical and the horizontal …"[1]

In other words, on a scope you are controlling the voltage per division and the time per division. The divisions referred to are the vertical and horizontal marks that make a checkerboard on the screen.

The next most important control is the capture mode, whether you are seeing a DC or an AC signal. Unfortunately, this control is usually somewhat hidden. This control is important because it can affect the way a signal looks on the screen. (Just take a 0 to 5 V logic signal and read it with your scope in AC mode and you will see what I mean.) In AC mode, the inputs are connected via a series capacitor to the guts of the scope. This removes any DC offset the signal might have. In DC mode, the voltage level of the signal relative to the ground lead of the scope is maintained.

The oscilloscope is, in my opinion, the single most useful tool an electrical engineer can have. That said (imagine a big sigh here), I've seen a lot of engineers chase down blind alleys because they misread their scopes. Correlating these two facts indicate that it is very important to know how to set up your scope.

First, a word of caution: Never trust the auto setup on a scope. Let me repeat: Never trust the auto setup on a scope.[2] Make sure you know what you are looking for. This is even more important than auto setups on meters because of what the scope might do.

---

[1] It is funnier if you think it in the same deadpan voice that the old TV show used. For those engineers who are too young to have any idea what I am talking about, you'd better Google *Outer Limits*.

[2] Note that I didn't say don't use it, I said don't trust it. You can use it if you have an idea of what you are looking for and can tell what the scope set itself to, to see if it is correct. It can save time if you use it carefully, but if you have any doubt at all, set it up manually.

For example, say you want to measure a 5 V signal that switches to ground when you press a button. You hook up the scope, press auto set, and then press the button. The most likely scenario in this case is the scope sees a 5 V DC signal and starts hunting for some frequency to look at. So it zooms in until you see a 10 mV AC ripple from the power supply at 60 Hz. Now you have a scope set to 10 mV per division vertically and 10 ms per division horizontally in AC mode. Remember, you were trying to set a 5 V DC switch to 0 when you press a button. The auto set totally missed what you were looking for. You probably won't even see the switch action at this setting and, to top it off, there will be a 60 Hz ripple on the screen to confuse you.

This is the most common mistake I have seen. An engineer hooks up a scope to the misbehaving circuit, hits auto setup, the scope zooms in on an irrelevant signal, the engineer, thinking "Aha, I have found the glitch!" spends the rest of the day chasing something that doesn't matter.

Having an idea of what you are looking for is an equally important rule for setting up a scope. Ask yourself how long the signal will last. What voltage levels do you expect? Start with those settings on your scope. Now, once you are capturing what you expect, zoom in on the details to look for those pesky glitches. Say, for example, you suspect a switch bounce on our earlier example. Start by capturing the signal at 5 V and 500 ms per division. After all, you are pressing this button—just how fast are you? Once you can reliably catch this signal, start working your way in; go to 2 V or maybe 1 V per division to increase vertical resolution. Then start working on the time base. Decrease the time per division while periodically checking the signal you are watching. This way you drive the scope to look at the signal you want to see. If you let the scope do the setup, it is kind of like being kidnapped and driven around blindfolded. When you take the blindfold off, you don't know where you are. You will be lost, confused, and disoriented and that can lead to wrong assumptions. If you are the driver, on the other hand, you know how you got there and have a better idea of what is going on.

So setup is important. Here are some other general things you should know:

Ask yourself, "Is the signal really there?" Why? Because it is possible that the scope with its high impedance is picking up noise that really isn't affecting what you are looking for. Try this: Disconnect the leads. Is the signal still there? If it is, that is a good sign that you are dealing with a radiated noise that might not even affect what you are looking at. If you are working with high-power circuits and switch-mode supplies, there will be all sorts of artifacts that really don't affect anything but that pick up nicely on the antenna of a scope lead.

Make sure you hook up all your ground leads (even though on most scopes they are tied together internally). The reason to do this is because small currents flowing back through your scope ground can lead to incorrect results. You might even think you have discovered free energy.[3]

On most scopes the ground lead is connected to the earth ground of the scope (for safety reasons), which can be disastrous when looking at certain signals that may reference to a different point. You can get currents through the ground leg that throw off your reading at best and blow stuff up at worst. If this is happening, get an isolated scope.

Just like a meter, high-impedance circuits can be affected by the scope leads. Have you ever had a problem go away as soon as you clipped the scope on? Try a 10 Meg resistor or 100 pf cap across the same connections. It is a good bet that will fix the problem (in case you were wondering about where those values come from, they approximate the impedance of most scope leads).

When all else fails, swallow your pride and read the manual. Yes, I know it's hard, but the destructions[4] usually give you insight into setting up the scope so that you see what you want.

Scopes these days have myriad features: cool glitch captures, colored screens (a personal favorite of mine), magnifications, auto setups (yeah, those too), and much more. The point here is to get the basic setup right so that when you use those other features, you have an idea of what is going on. Remember, getting what you want out of the scope is up to you, at least until they get that mind-reading function working.

Logic Analyzers

A logic analyzer is similar to an oscilloscope in that it displays a signal over a time base. It differs in two main aspects: The first is that it displays only logic levels; the second is that it has many more channels.

Think of a logic analyzer as a digital-only oscilloscope. It is not going to show you signals between a logic high or low. There are logic analyzers with a couple of scope channels built in to get around this limitation, but if you don't have one of those, make sure that you understand you are seeing the logic level closest to the signal you are reading. If the level the analyzer considers a high or

[3]This is a whole other topic for a whole other book.
[4]Or instructions, depending on how you look at it.

low differs from the level of your circuit, this could lead to confusion. If you suspect that the logic signals are not reaching the required voltages, make sure you check it with a scope.

The best feature of a logic analyzer is the fact that it has so many channels. This becomes very useful when you are trying to observe all eight or more lines on a data bus at the same time. It's kinda hard to look at eight things at once with only a couple of channels.

This feature, like all the others, is easy to set up wrong if you have no idea what you are looking for. Don't just set it up blindly—have an idea of the time base needed to find what you are looking for. Also, remember that it is designed to display logic signals, possibly masking signal levels that you might not expect.

These days, with their digital storage capabilities, scopes are closer than ever before to logic analyzers, and the fact that many analyzers have some scope-like capabilities make them more scope-like than their predecessors. If forced to categorize, I would say that a scope is a more general tool that can be applied in nearly any situation except the one where you need to see a whole bunch of channels at once, and in that case the logic analyzer is definitely the tool of choice.

Remember that the basic rule of thumb with this tool, as with all others, is to have an idea of what you are looking for. If you do so, you will find this an effective tool to have at your disposal.

Thumb Rules

- Always have an idea of what you are looking for.
- Don't trust auto setups.
- Is the signal really there? Unhook the leads and see if you still pick it up.
- Hook up all the ground leads.
- The higher the impedance of the circuit, the easier it is to disturb with measuring tools.
- Read the manual!
- And one last time, don't trust auto setups.

SIMULATORS

First, let me make a statement: Simulators are great tools (here it comes), *but* too often I see a major mistake made with a simulator. The engineer fires up the simulator, tries out his or her idea, gets it all designed, then proceeds to build a real circuit, only to find the circuit does not work as the simulation did. Here is where the mistake comes in: All too often the engineer spends all his or her time trying to figure out why the circuit isn't working right while implicitly trusting the simulator to spit out the correct answer. For some reason as soon as the circuit is modeled on a computer, it seems to be an engineer's nature to trust the result on the simulator without question. Doing so almost invariably leads to immense frustration and confusion. You should take this adage to heart: *The real world isn't wrong; your simulation is*. It is always true. If the results don't match, something in your simulation does not actually represent what is on the prototype in the lab. The simulation is a representation of the real world, not the other way around.

What Is Real?

This is not to say that the circuit on the bench is what you want it to be. It very well could have a mistake in it that is not in your simulation. However, that doesn't change the fact that the simulation is not truly modeling your design. I have found that if you take the perspective of always questioning the simulation, two things happen. First, you gain an intuitive understanding of the way different components affect your circuit. As you fiddle with the simulation, trying to get it to match the real world, you begin to grasp how large an effect this or that component has. Second, you learn about the limitations of real-world components—something that just studying math and formulas will not give you. Take, for example, a 10 µf electrolytic capacitor in the circuit shown in Figure 5.3.

According to all the formulas you have learned, this should pass all the high frequencies above 1/RC you would ever want. Just about every simulator you find will do so, but hook this circuit up to a signal generator and you will find that, as you get up to the higher frequencies, it doesn't work as well as the math says it should. The math isn't wrong; it's just that the component isn't perfect.

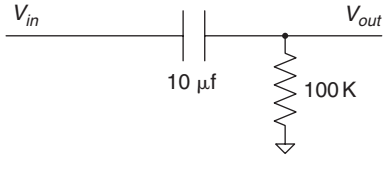


FIGURE 5.3
RC high-pass filter.

Some simulators will allow you to create equivalent circuits to more accurately represent a given component. Remember, though, that doesn't negate the need for you as an engineer to understand the limitations of the components. You really need to have an idea of what is going on or the simulation can lead you down a fruitless path. The skill of estimation is immensely important when using a simulator. Skip back to Chapter 1 if you need to brush up on your hand grenade skills.[5]

A Powerful Tool

Now that I've finished bashing simulators for not dealing well with imperfect components, let me say that, ironically, they are potentially the best tool you have to create a design that handles imperfect components well.

Once you truly understand the variability that can occur in the parts you are using and create an accurate model of what they do, you can do something with a simulator that you cannot do easily with actual parts: You can build thousands of pieces of your design in cyberspace, with each part varying a little from its nominal values. You can swing the tolerances to their extremes with the click of a mouse, saving a hunt through a drawer for that part that is on the low end of spec. If used correctly, a simulator is probably the best tool you have to make your design handle the inherent variability in components.

Develop Your Intuition

One of the best things you can do with a simulator is to use it to develop your intuitive understanding of basic components. Every engineer should simulate the transient response of the basic RC, RL, and RLC circuit. Try changing the values of the parts just to see what happens.

If you start modeling simple circuits and getting confidence in making the model accurate, you will be much more successful as you create more complex simulations. It's not unlike learning to play the guitar; you don't just sit down and rip out a lick Eddie Van Halen would be proud of. You need to be able to handle the basic chords first. You should learn to "play" a simulator the same way.

Even though it is easy, don't put together your whole design in the simulator the first time and press go. If you do, I can nearly guarantee you will get

[5]Admittedly, that section is personally my most favorite passage in this whole book!

confused by the results and they will probably be wrong as well. Break your circuit down into simpler pieces, ones that you can intuitively understand, and simulate those parts first. Eat the elephant one bite at a time.[6] When you are sure your model represents the real world accurately enough[7] for the problem at hand, start knitting those pieces together and see what happens.

One word of warning: Playing around with a simulator can be very time consuming.[8] Don't get so caught up in doing the simulation that you never get around to building an actual circuit. In fact, if you are unsure as to how the circuit will really work, go build it up in the lab and see. When it comes to tolerance analysis, you should already have a real circuit running in the lab when you start simulating. Get the circuit working with nominal values before you start investigating what component variance will do. Simulation should go hand in hand with lab work.

Thumb Rules

- The real world isn't wrong; your simulation is.
- Gain confidence that your model accurately represents your design.
- Use estimation to double-check your simulation (a couple of more -tions and this could be fun to say!).
- Model basic circuits to develop your intuitive understanding of the basic components.
- Break the model down into pieces that are simple enough to check for accuracy. Then add the models together.
- Simulation goes hand in hand with lab work.
- When setting up your tools, have an idea of what you are looking for. How fast is the signal? What voltage level do you expect it to be at? et cetera!

[6] See Chapter 1 way back at the beginning for the elephant reference.
[7] Remember that accuracy is relative. If you don't need to know the answer to four decimal places, don't waste time trying to get that close.
[8] Not unlike research on the Internet. Well, maybe that only holds true for a "sparky."

SOLDERING IRONS

I was passing by the lab one day when I saw one of my technicians looking over the shoulder of one of the engineers who was doing a less than spectacular job of soldering components on a PCB.[9] He had but one comment. He said, "What we have here is an engineer trying to do a technician's job." Then he sat down and proceeded to do a most excellent job of putting the board together.

On the chance that you might not have a skilled tech at your disposal, and due to the fact that I believe that the more you know about how the product you are designing goes together, the better designer you will be, here we will go over the basics of soldering.

The Basic 4

Making good solder joints requires four things: cleanliness, solder, flux, and heat.

First, the parts need to be clean and dry. If the pads are corroded, often a little rubbing alcohol will clean them nicely.

Second, you need solder. Solder is a mixture of lead and tin[10] with a melting point around 100° to 200°C, depending on the alloy used. When applied properly, solder will provide an electrical and mechanical connection between the part and the PCB. Although it is a mechanical connection, remember that it is not a particularly strong mechanical connection.

Third, you need flux. When hand soldering, this is often inside the solder wire in the hollow core. What is flux, you ask? Flux is a chemical that cleans when you heat it up, preparing the joint so that the solder will stick well. In some cases the flux is applied before the solder, such as before it goes over a solder wave or into a solder bath. Flux is also called resin.

Last, you need heat. Heat brings it all together. The solder will flow to where the heat is. This means that you need to get the leads of the part heated to make sure the solder flows. In prototyping, the typical way you get heat to the part is with a soldering iron. Some other ways are hot air pencils and reflow

[9] One engineer I worked with developed his soldering skills putting stained-glass windows together as a part-time job when he was a student. After quizzing him on the technique, I recommended he do exactly the opposite of what he learned making windows!

[10] These days you will be treated to stuff called RHOS-compliant solder, which uses different stuff inside and can be a bit more finicky to use. I suggest a little higher heat on the iron, but take care not to damage your part with too much heat for too long!

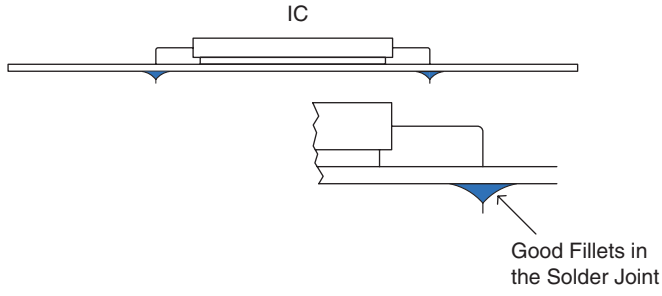


FIGURE 5.4
Good solder joint.

ovens, but the same thing applies. Heat makes the solder adhere to the pad and the lead of the part. When all is said and done, a good solder joint looks like the one in Figure 5.4.

Solder Goobers

Of these four items, the one that usually causes problems is the application of heat, particularly when you are using a soldering iron. Parts and PCBs are both sensitive to heat. The parts can be damaged by too much heat, and the PCB pads are adhered to the PCB with glue that has a lower melting point than solder.[11] Too much heat for too long can be bad. Parts can be damaged and pads or traces can be lifted (when the glue is melted).

The flip side is that not enough heat will lead to failures. One of these failures is called the *cold solder joint*. This happens when you do not get enough heat to both parts being joined. When this happens, solder will adhere to one part and not the other. The part that did not get enough heat will not get a good connection. That is why it is said to be a *cold* joint. It looks like Figure 5.5.

A cold solder joint is the most common failure of using a soldering iron. You get going a bit too fast and don't leave heat on the joint long enough, or you only touch the iron to the pad and don't get it on the lead of the part. A good rule of thumb when soldering by hand is to place the tip of the iron on the joint, count "one Mississippi," and then apply the solder, wait a moment, and remove the iron.

[11] It is actually intended to be this way because during soldering the copper traces will expand (due to heat) at a different rate than the PCB substrate. If the glue is melted, this keeps the trace from deforming.

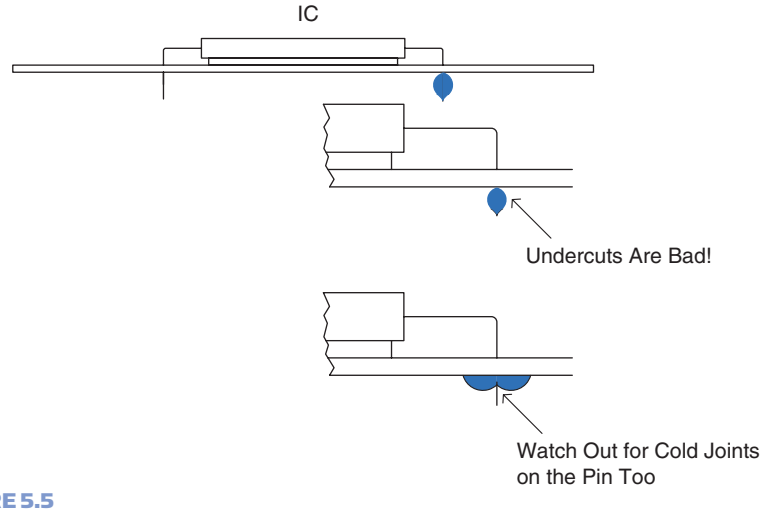


FIGURE 5.5
Cold solder joint.

There are two other things you need to do to keep your soldering iron working right. One is to make sure the tip is tinned. If an iron is left on for some time, the solder and resin on the tip will evaporate, leaving a dry tip. A dry tip will not conduct heat to the parts you touch as well as a tip with solder on it. Applying a little solder to the tip before using it is called *tinning*. (You can also tin wires to make them easier to solder to a connection.) If you are having a problem getting heat to a part, try adding a little solder to help conduct the heat.

The second thing you need to do is clean the tip of the iron often. Any decent soldering iron will have a sponge in a tray with water. Wiping the tip on it will effectively clean it. Cleaning the tip keeps the buildup of excess flux from interfering with the soldering process. A word of caution: Don't soak the sponge with too much water, and don't rub the iron on the sponge excessively. Too much water or rubbing it too long will cause the tip to cool down too much, affecting the next joint you need to apply solder to. Don't forget to tin the tip before going onto the next joint.

SMT Specifics

In today's world you will likely be treated to the fun of surface-mount components. Though seemingly impossible to do by hand, they are not as bad as they might seem. I recommend you use solder paste; you basically paint the leads and the pads with the paste. Then with a nicely timed sweep of the solder iron you can get all the leads connected with some very pretty joints. It takes some

practice, so be prepared to go through a few parts and possibly PCBs the first time you attempt it. Too much paste and you will get shorts across the leads, too little and you get no connection. When you get the right amount and the right timing of the iron, the solder flows to the right places and does just what you want.[12]

Desoldering

Unless you never make a mistake, at some time in your career you will need to remove a part that has been soldered to a PCB. Desoldering can be a frustrating experience. It is during desoldering that you are most likely to lift a pad from a PCB, burn your fingers, and possibly cut loose with a few expletives. Hopefully, I can share some hints to keep the air from turning blue when desoldering is required of you. We will also discuss the three main tools for lifting solder: solder tape, hand pumps, and desoldering stations.

Hint 1 Sacrifice the component if possible. If you do not have to salvage the part you are taking off, clipping the leads (so that you are not trying to remove a 40-pin part all at once) is a great help. Cut all the leads and deal with one pin at a time. Once we had a situation where we needed the 40-pin part but not the PCB. What was our solution? Take the board down to the shops and hit the back of the PCB with a quick burst from a blowtorch while yanking the part off of the other side with a pair of pliers. It worked like a charm, and the burnt PCB made for a great joke on management later!

Hint 2 Add solder to the part. Adding solder can help you conduct heat to the joint you are trying to dismantle. The trick to getting the part off is to get heat quickly to all the places you need to. For example, you might need to remove a radial electrolytic capacitor. On this part both leads are close together. You can actually create a solder bridge between the leads and get heat to both leads at the same time and quickly pull the part off.

Hint 3 Get the part and pin off before you worry about getting the solder off. Apply heat, yank the part, then come back and get solder out of the hole. Often when you are trying to get the solder completely off before taking the part off, you will find that a small piece of solder still holds

[12] If you ever get a chance, watch an SMT reflow oven do its magic. The solder will actually have enough surface tension to align the part when it is melted. I think it is pretty cool to watch. (Yes, my wife thinks I'm weird to think this is cool, but if you bought this book, chances are you will think it's cool, too.)

the lead to the side of the *via*. Trouble is that it is such a small piece of solder that it is difficult to heat it up to get the lead loose. Apply hint 2 and try again.

TAPE

Solder tape is a copper braid. Copper, being a great conductor of heat, will wick the solder into the braid when heated up. It is important to apply the heat to the braid and then press the braid on the solder. If you just try to stick the braid in the molten solder without heating it up, the solder will just sit there. Remember, solder flows to heat.

Also note that the braid is made of copper, and copper can tarnish. Once it has tarnished, solder will have a hard time sticking to it, so old solder tape is pretty much useless. New tape works well, though, and is cheap and convenient to use.

HAND PUMP

My own tool of choice, the hand pump, is easy to use, relatively inexpensive, and easy to maintain. When using the hand pump, you press down the plunger, heat up the solder you are trying to remove, press the button, and *thwoop*, in goes the solder like a spaghetti noodle. Make sure you leave the iron on long enough for the solder to become molten clear through the *via*. You might need to apply hint #2 to help things out. The biggest downside to the solder pump is the sore thumb you are going to get if you need to do a lot of desoldering.

DESOLDERING STATION

If you need to do a lot of desoldering and you have some cash to spend, this is a tool you need. The desoldering station is a powered version of the hand pump. The iron is integrated into the tip, where a vacuum is applied to suck out the solder. Generally you need to maintain these tools regularly. The tips can wear due to the corrosive nature of the solder removal. They can get plugged easily when they're not used properly. Always suck to the side, not straight up. The molten solder has weight, and trying to move that to the side is easier than trying to lift it straight up. Keep sucking for a couple of seconds after the joint is clear to make sure that the molten solder gets all the way into the receptacle in the gun so that it doesn't solidify midway through the nozzle.

Properly maintained, this is the quickest and easiest way to get solder off a PCB. It is also possible to get a part off with the pin still in place. This is done by using a small circular motion to get the pin out of contact with the *via* as

you are sucking the solder. However, it is still easier to sacrifice the part if that is possible.

Thumb Rules

- Solder goes where the heat is.
- Solder goes where the heat is.
- And if you didn't get it this time, remember: Solder goes where the heat is!
- Prevent cold solder joints by counting 1 second while applying heat.
- Make sure that you tin the iron before using it.
- Use just the right amount of solder paste with SMT parts and just the right amount of time.
- Clean the tip often.
- Practice makes perfect.
- When desoldering, sacrifice the part if possible.
- Add solder to promote heat flow.
- Get the part and pin off before worrying about getting the solder out of the hole.
- A small circular motion with a desoldering station tip will help clear the solder from the lead and the via.

PEOPLE TOOLS

When I entered the professional realm for the first time, I had an experience that I still remember. I got this call from the receptionist. She said, "So and so is here to see you. He wants to know if you can have lunch with him." Of course I'm thinking, "Hey, free food, but who is this guy who seems to be my instant friend?" Thus it was over nacho chips and *arroz con pollo* that I tumbled headlong into my first experience in the world of reps, distributors, and FAEs.

Lunch was good. I had no problem figuring out what to order from the menu, but getting to understand the roles of these three people took more than a few tacos. It can be a bit confusing as to who does what and what that means to the average "Dilbert" out there, so I figured it wouldn't hurt to give you some idea of what these guys do and how they can help you.

First, all these people have some relationship to the company that makes the product you need, whether an IC, transistor, micro, or whatever. When I say *company* in this case, I am referring to the company with the product to sell, not the company you work for.

The Company

The company selling the widget you are interested in employs several layers of people to get its product in front of you and sold to you. It also has internal salespeople and managers you might get to know if you work closely with them. How closely you work with them often depends on the amount of business or potential business you represent for them.

If you can get to know someone on the inside, it is never a bad thing. These guys are more accessible now than ever, and if you need to know how a part will act in some weird situation, talking to the guy who actually designed the part is definitely the best option.

The Rep

One layer removed from the company, you will find the rep.[13] This is the person who represents the company making the part it wants to sell to you. He or she does not usually draw a salary directly from the company. Reps are paid by the rep firm that represents the company. The rep and/or her firm are typically rewarded for her efforts with some percentage of the sales she makes, usually 1 to 5%, depending on volumes and other complicated formulas designed to cost the company making the product the least amount of money yet drive sales as much as possible.

The rep will work with the distributor in scheduling parts, getting you samples, and other such stuff. She is very interested in you using the company's widget, and it is unlikely she will offer an alternate solution since she is monetarily tied to the company's widget. Reps typically are not allowed to represent competing firms. My experience is that reps for the Dilbertesque products we are talking about often have an engineering background of some type.[14]

The Distributor

"What is the difference between the distributor and the rep?" I asked one of these guys once. "About 15%," I was told.

---

[13] In case you didn't know, *rep* is short for *representative*.

[14] If they don't, I know where they can get a great book that makes it easy to understand the world of "sparkies"!

Distributors will stock parts and mark up the cost to cover the money they have expended. They tend to make about 20% on a given part, but that is just a ballpark figure. The actual number can be all over the place, depending on the particular business agreements. Some manufacturers force distributors to specific margins if they carry their parts.

One of the biggest distributors out there watches market trends constantly and looks to buy stuff that is likely to become rare yet needed in the *future*.[15] Then they go out and buy a whole bunch of said item, sit on them for a bit, and sell them at a profit later.

Sometimes companies use exclusive distributors. Some use multiple channels. In the case of multiple distributors, whoever is the first to register[16] a part for a particular application gets a lower price on the part than any of the other distributors. This is designed to reward the distributors for getting out and getting more business.

The biggest advantage of a distributor is supply chain management. By buffering stock for you, they can help handle ups and downs in order sizes, shortening lead times when orders go up unexpectedly.

They are less likely to be tied to a particular manufacturer of a part, and they often carry multiple solutions to a given problem. They will tend to lead you to the part that will solve your problem and be the most profitable for them.

Distributors are less likely to have an engineering background. Hopefully many of them will buy this book and not be upset that I disclosed so much about this seemingly secret world.

FAEs

Working for distributors, reps, or even the company, you will sooner or later run into the *field application engineer* (FAE). The FAE plays several roles. He is the main guy who helps you get the part to work. He or she also looks at your application and often might suggest parts that would be a good fit. Lastly, FAEs often act as translators between you and the distributor. As you might well know, it can be difficult to understand an EE when he or she gets into technical details.

---

[15] Slight pun intended; if you get it, I don't need to explain it, and if you don't it is no big deal.

[16] In registering a part, the distributor basically calls dibs on showing you the part first.

For many a Dilbert, the FAE has the perfect job. The FAE gets to come up with all these solutions but has no responsibility for actually making it work at the end of the day. There have been days I have dreamed of being an FAE for just that reason. However, the flip side is FAEs also rarely see the finished product and miss out on the satisfaction of the "being late and over budget, but whew! It's finally done" feeling.

FAEs often go to a lot of training sessions with the company to understand how the part works. They usually know or can contact the engineers in the company to help answer questions. More and more, in an effort to sell their parts, companies are not only developing new parts but also creating applications for those parts. FAEs and company engineers are often tasked with coming up with cool little application demos and the like that show you how great the part is. Remember, though, for them it is like lab back in school—they only need it to work that once when they are showing it. Production runs can be a whole other matter, so do your homework and test an FAE design thoroughly before you commit to it on a full production run.

Design Wins

When the distributor registers a part with the rep and thus the company and then the part actually gets used in the design, it is called a *design win*. This is a common term that's heard often over appetizers. Odd how engineer types use words that make sense when you think about it, isn't it?

Remember, whoever registers the parts gets a discount on that item. That usually makes their price hard to beat compared to other distributors. I say *usually* because I have seen a lower price quoted from an unregistered distributor, albeit rarely.

Another thing that can happen is that registration can be moved. If you really don't like working with the distributor and you are a big enough customer to the company, they can move the registration to a distributor you prefer. However, this is rare and usually done as a last resort to appease the customer because if it happened too much, all incentive to get their part in the door first dries up.

Going Direct

Depending on the size of your orders, one thing that you might consider is going direct. This means that you will buy parts directly from the company, skipping distribution. The goal is to get a lower price.

The cons to this approach are several. The company will usually have minimum orders, lead times, and terms that are less favorable than working with a distributor, and if you are a little guy (order wise) they probably won't even consider it.

Before you do this, consider the options carefully, because you will be removing a piece of the support structure that you use in the design and supply management of the part.

There are companies that will not even allow you to go direct; they have a policy of distribution only. However, I happen to know that they also dictate to those distributors what the final price will be to keep them competitive in the marketplace.

To Sum It Up

There are several legs to the stool of getting parts to you. Each leg wants his piece of the pie and has services to provide to justify their cut. Knowing who does what will enable you to better work with these people tools to succeed at your job.

In my experience, the more successful reps, distributors, FAEs, and the like will visit you often enough to know what you are working on and keep you in mind as they see new technologies and ideas that you can use. They will have suggestions and solutions and, yes, they might even buy you lunch once in a while.

Thumb Rules

- The rep works for the company under a contract.
- The distributor works as an independent.
- The FAE knows how the stuff works.
- The company wants to sell you a cool widget.
- Sometimes you can get the company and the FAE to help do some of the design work.
- All these guys can help you find parts and get quotes; they work together to provide the best service they can.

CHAPTER 6

Troubleshooting

The perfect design approach doesn't guarantee that everything will go flawlessly in production. One engineer I worked with was fond of saying, "I'm not happy till ten thousand pieces have gone down the line okay." I've seen tolerance problems appear in production after half a million pieces have been run. It is very difficult to predict and prevent something like that.

The fact is, the more thoroughly you try to analyze a design, the longer you will be waiting to produce it, but if a product never gets into production, no one gets paid (except in government work, of course). So, a balance needs to be struck among design analysis, testing, and production runs. This being the case, you will likely be faced with trying to determine a production problem at some point in your career. How will you go about it? What approach will you take? Hopefully this chapter will give you some ideas.

GETTING READY FOR THE HUNT

As we discuss the topic of finding trouble and shooting it, I will often refer to my own experiences. I am sure you will have unique and often completely different results. The idea here isn't to tell you what the problem is in your design but to give you some guidelines you can use to troubleshoot the problem yourself.

Shotgun Wedding

We will get into the interesting section header a bit later. Though it might seem a little out of order, I think we should cover one item first because it is so

important to the rest of the process. The first rule of thumb is: Don't discount a theory (no matter how obvious or ridiculous it might seem). Try to prove it right or wrong by experiment and then move on to the next idea. Too often you will be carrying an assumption that you won't even realize can lead you to a wrong conclusion. It is vitally important that you have a process to check and validate a theory. Without that, you will be forever jumping from one idea to another without ever coming to a conclusion. With that said, let's move on.

When it comes to troubleshooting methods, I group them into two common categories:

Scientific method: Do what any good detective would do: Look for all the clues you have been given and deduce what the problem might be based on experience and knowledge.

- *Advantage.* Eventually you will identify the problem.
- *Disadvantage.* It takes a lot of patience and time.

Shotgun method: Take a shot at as many possibilities as you can and hope you get a hit. Sometimes you get lucky and you solve the problem fast.

- *Advantage.* If you are lucky, you will solve the problem fast.
- *Disadvantage.* If you are not lucky, you will chase around in circles forever.

Although both these methods have their place, what I propose so subtly in my title is a third approach to troubleshooting. I call it the Scientific Shotgun Method—a marriage, if you will, of the shotgun method and the scientific method. Start like this:

When a problem first comes to your attention, take a shot at as many possibilities as you can. Write down all the things you think might be causing it. Use your intuition as well as your experience in this exercise. Speaking metaphorically, get out the shotgun, take aim, and fire. Then let the scientific method kick in and figure out a way to evaluate each of your possibilities to prove or disprove them, and have at it.

When employing the scientific shotgun method, based on my experience, results like these are typical: 70% of the time it will be something stupid that the shotgun method catches easily and quickly. For example, using an old software version or if a component wasn't stuffed or a fuse was burned out, 20% of the time something more subtle will be found that takes some trial and error and requires new data to be found and evaluated till the problem is solved. About 10% of the time the solution takes a while longer but eventually is

found by repetitive applications of both methods. Often the shotgun approach will open up new areas of research that scientifically lead to the resolution. On the aggregate, problems are typically solved quickly with a minimum of running in circles when the *scientific shotgun* approach is used. (Did you ever think you would see those two words together as something meaningful?) This is a real boon in a consumer product world when shipping that new design on time is all-important.

You Too Can Learn to Shoot Trouble

Have you ever seen an engineer having immense difficulty in diagnosing the cause of a problem when a lowly tech stops by and identifies the bad part right away? Or maybe you've seen a tech struggle for days only to have the engineer take one look at the schematic and say, "There is your problem."

Some people have trouble with troubleshooting, and others just seem to have a knack for it. If you ask them to explain what they do to solve problems so quickly, they are often at a loss as to how they do it—they just do. Believing that you can learn anything even if it doesn't come naturally, I have distilled down into Thumb Rules some of the things that those guys with the knack do.

Simple Things First

After you have made the list of things that could go wrong, start with the simple things first. My father recounted an experience to me when I was younger that really stuck with me. He basically rewired an entire car looking for an electrical problem. To his dismay it turned out to be a bad fuse. Looking at it, it appeared okay, but when measured, it was open. This might seem like a "duh" moment to the outside observer, but it is an easy trap to fall into. The way to avoid it is to check out the simple things first. Does the chip have power at its pins? (Not just to the board.) Is the oscillator running? Walk your way through the simple stuff, avoiding assumptions and checking everything for simple failure first.

Look Outside Your Specialty

It's hard to make a blanket statement about what is likely to fail, since there are often many small clues to a particular problem. To further complicate things, it is often a combination of more than one factor that's causing the problem.

It is human nature to focus on what you know; everything else seems somewhat magical after all. Good troubleshooters are often good generalists. They know a little bit about everything and use that to connect the cause to the

effect. They always want to know why this is that and what does that thing do, and so on.[1]

Sometimes there could be seemingly insignificant clues. One time early in my career we had a problem with some displays we were producing. A percentage of them were failing and I was assigned to find out why. When I took the unit apart, it would function correctly. When I put it back together, it would fail again. I looked for hours trying to find problems with pinched wires and cold solder joints, to no avail. So I sat there and stared at the PCB for a while.

As I did, I noticed two small marks on a resistor; I wondered where they had come from. After some examination, I discovered a screw head that would contact this resistor when the PCB was installed. It turned out that the screw head would short across the resistor when the PCB was installed, making the part fail. When I removed the screw, the console worked correctly after assembly. Don't be afraid to look outside what you know for the cause of the problem.

Don't Ignore Anything

Try to keep track of all the clues to a particular problem. Keep a list of symptoms and clues that you can refer to in your deductions. Don't ignore anything, since one fact might connect with another to point you in the right direction. Here is a case in point.

During testing of a control circuit my engineering team had designed, we had been experiencing random unexplainable problems. The test engineer made the statement that these problems seemed to have started when we began using surface-mount PCB designs. We were completely at a loss as to any connection between this and the problems. Then I remembered, when looking at one of the circuit boards, I had noticed some small black fibers that appeared to have dusted the PCB. The test engineer initially dismissed this as small bits of plastic that accumulated in the environment of the circuit during use (this made sense because there was a moving belt made of plastic that could leave these bits as it wore down). He was sure it wouldn't make a difference.

However, we knew there were points on the PCB that, if they were shorted by even a few mega-ohms, could make the circuit repeat the problem that we were seeing. Connecting that with the fact that the surface-mount components would have closer spacing made such a short more likely. I insisted that we determine conclusively whether these fibers were conductive.

---

[1]You know—the kids who drive moms nutty because they are always asking questions. I know the personality type well. I have five of them!

The first thing we did was collect a sample of this "dust" and bring it near a magnet (on the presumption that if it is ferrous it is likely conductive). We were surprised at how much ferrous material was in these presumed plastic shavings. It reminded me of the classic physics experiment where you put metal filings on a piece of paper and then move a magnet underneath to see the field interactions. Once we protected the board from this contamination, the strange behavior stopped.

By not dismissing the obvious presence of these fibers, combined with the clue that it started when we went to an SMT design, we were able to make a connection that allowed us to solve the problem.

Which of These Things Is Not Like the Other?

Did you watch *Sesame Street* as a kid? One of my favorite segments was "Which of These Things Is Not Like the Other?" You were taught to identify similarities and then point out the one that just didn't seem to fit. This is a very important troubleshooting skill. All the good skills work in more than just the "world of sparkies." They can be applied in just about any problem hunt. Here is another case from the Archives of Darren.

Years ago our fridge stopped dispensing water. I figured I should just tear into it and take a look.[2] After all, the water valve was controlled by a solenoid. That was close enough to electricity for me. There were two valves, one for water and one for ice. I tore these valves apart and noticed some wear on a rubber washer inside the water valve. The solenoid pressed this washer against a hole in the valve. Little bits and pieces of rubber were falling off because it was so worn. This became especially important after I looked at the ice solenoid (it was operating correctly) and the rubber washer on that one didn't show wear.

It just didn't seem right for the washer to be falling apart like that. It just didn't fit. So I replaced the rubber washer. Put it all back together and, voilà, it worked great. The skill in this case was looking for something that simply didn't seem right. Sometimes you can figure this out by asking yourself, "Would I have designed the washer to fall apart?" The obvious answer in this case was no, so hence something was wrong with the washer.

---

[2]Early in our marriage my wife was dismayed at my willingness to tear into anything we own to try and fix it. After a few years and a few successes I think she came to appreciate my "knack," because now she seems to expect me to fix nearly anything and wonders at my inability when I fail to do so.

Estimation Revisited

Sometimes it seems we spend half our time designing a circuit and the other half trying to figure out why it isn't doing what we designed it to do.

Back in Chapter 1 we learned to develop an intuitive understanding of basic components. An important part of this process was developing the skill of estimation, to get an idea if the circuit is even close to where it is supposed to be.

Estimation plays an important role in troubleshooting as well. If you are good at estimation, your intuition will be correct and will point you down the right path to solving the problem. Combine that skill with the power of the modern-day calculator, and even a circuit simulator as we talked about in Chapter 5, and you have a powerful toolset to diagnose the root cause of a problem.

Can You Break It Again?

This is a simple rule that is often overlooked. Once you have found and corrected the problem, can you break it again?

That is, can you remove the fix and see the circuit act up again, doing whatever it did before? Often, especially with problems that are difficult to repeat, an engineer will apply a fix, have the problem seemingly go away, and figure he is good to go. However, if the problem is a bit temperamental, meaning it doesn't always show up when you want it to, you might just coincidentally apply the fix when it went away on its own. In my experience this can happen quite often, so I recommend breaking it again to see if you are really fixing it or not.

It's no fun to think you have fixed a problem only to fire up the production line and shut it down again when the problem reappears. You can also spend a lot of money applying fixes that are not really needed. So please break it and fix it several times before you are sure you have the problem solved.

Root Cause

A good troubleshooter will methodically trace an offending signal back to its source. As he does so, he will question each component in the circuit as to whether it is operating correctly. He will ask himself things like "Does the output signal of this op-amp agree with the signals that are on the input pins?"

This is why the really good engineers seem to always be muttering to themselves. They aren't schizophrenic, they just ask themselves a lot of questions. (Okay, maybe they are schizo, but trust me, it's in a good way.)

Eventually you will find the problem's root cause—the component that isn't doing what it is supposed to—and then you can figure out why and get it corrected.

Categorize the Problem

Good troubleshooters will separate the problems into various buckets and use an approach that works best for the type of problem suspected.

Design problem: This is the most common mistake and the easiest to find, since it is generally repeatable and consistent.

- *Approach.* Since it is repeatable, keep it misbehaving while you use tools (scopes, meters, etc.) to trace down the problem. Make sure you get to the root cause.

Tolerance problem: Really a design problem, but I give it a special category because this is typically inconsistent and difficult to repeat. Environmental effects commonly aggravate this type of problem.

- *Approach.* You will need to repeat the environment that caused it if possible. Here is also a good place to run simulations where you can vary the tolerance of the parts you suspect and see what happens.

EMI problem: This can also be difficult to repeat. Who knows when EMI is going to hit? It will often trip up the most competent engineers.

- *Approach.* This one is so much fun I have dedicated a whole discourse to it, coming up next!

Software problem: So many products today use some type of software or firmware. I have seen software exhibit all the preceding symptoms and be used to correct some of these problems, even though it was really a hardware issue. This topic gets its own category for that reason.[3]

- *Approach.* Give up, go home. No, not really, but it is a fact that these can be a bugger (pun intended) to figure out in a reasonable amount of time. Combine that with software engineers' natural fear of oscilloscopes[4] and you can see you are in for a treat when diagnosing a software problem. The longer this paragraph gets, the more I think it needs its own discussion, so I put one in. We'll get to it in a bit (pun intended again!).

---

[3]Here is a metaphorical question that will drive your code jockeys nuts: If you can fix a hardware problem with software, was it really a software problem in the first place?

[4]Maybe I am wrong, but it seems like I am constantly reminding the software engineers to get out a scope and have a look at the signals they are making happen with their code.

Go Shoot Some Trouble

Now that you have some basic skills, put them to the test: Take aim and blow that trouble out of the water! As one last idea, keep notes of what you are looking into and the conclusions you are drawing. This is especially important if what you are looking for is taking a while to find. It is also nice to have when you are creating your design guidelines. You can refer to these notes to know what not to do in the next design.

I know it sounds like those dastardly lab books you had to keep while in school, and it is, but remember, you aren't getting graded on them. Just keep the notes in a way that makes sense to you. Take some notes, get out there, and blast trouble away.

Thumb Rules

- Do not discount a theory outright; try to prove it right or wrong by experiment.
- Use the shotgun wedding approach to get to the root of the problem quickly.
- Start by checking the simple things first.
- Look outside your specialty.
- Don't ignore anything, and the corollary, don't assume anything.
- Look for what doesn't belong.
- Use estimation and intuition to lead you in the right direction.
- Dig for the root cause.
- Can you break it again?
- Categorize the problem and customize your approach.

GHOST IN THE MACHINE: EMI

Have you ever had a circuit or design do something you don't want it to and you just can't explain why it does it? Worse yet, it doesn't do it all the time, just when the planets are properly aligned. You might just have a circuit haunted by EMI's (pronounced *Emmy's*) ghost. Dealing with EMI is definitely a school-of-hard-knocks course. Here are a few "Cliffs Notes" for those of you who are about to enroll.

EMI stands for *electromagnetic interference* and boy, does it ever interfere! I remember one of my first bouts with this ghoul. We had recently completed a design of a display that worked great on the bench and even worked most of the time on the product. However, about 20% of the time when we turned the motor on the display would simply freak out. By an all-night process of trial and error, we finally stumbled across a solution to get production up and running again. Since then, I have learned a lot about how to pinpoint an EMI problem and resolve it. The things I point out here work well when combined with the troubleshooting techniques previously discussed.

Few engineers have ever dealt with EMI on anything other than a troubleshooting basis. Let's face it, we don't go looking for EMI, it does just fine finding us by itself! Let's start by getting a basic understanding of what EMI is.

What Is EMI?

EMI is basically an unwanted signal entering your circuit. It is still an electrical signal, it still obeys Ohm's Law, and, for all its supernatural behavior, it is still just a signal. This is good news! It means that you can exorcise these demons from your design because they still obey the laws of physics.

The Ways of the Ghost

First, how does EMI get into a circuit? There are only two ways: it's conducted or it's radiated. In the first case, the unwanted signal has to travel on a trace, wire, or other directly connected path into the area of disruption. In the second case, the signal propagates without wires. It is important to know how the signal is getting in because that affects the solution you will need to employ.

Conducted EMI

How do you know if it is conducted EMI? The easiest thing to do is disconnect everything part by part until the problem goes away. Case in point: We were hooking a computer up to a circuit board, both at the audio output of the sound card as well as the serial port. There was an annoying buzz in the speakers that changed tone in sync with the displays on the board. When I unplugged the serial connection, the buzz went away. We had what's known as a *ground loop*. This is a specific type of conducted EMI. I usually try to detect whether the problem is conducted EMI first, since this is the easiest to check. Don't overlook the connection to a wall outlet if the AC line powers your device. I once saw a design disrupted every time an overhead projector was plugged in.

Radiated EMI

The best way I have learned to determine radiated effects is to divide them into two camps: the *near-field effects* and the *RF effects*.

Near-field effects can be easily divided further into current and voltage disruptions. Consider this rule of thumb: Anything within a wavelength is near field and anything outside that range is RF. Inside the near-field range, magnetic fields induce current fluctuations into a circuit and electric fields produce voltage fluctuations.

Here is a simple test with a piece of equipment that you are likely to have on your bench. Take your oscilloscope probe and, with the ground dangling, move it near an AC outlet. Adjust the voltage range and quickly you will see a nice 60 Hz sine wave. This scope configuration is basically a dipole antenna and it responds well to electric fields.

But what about magnetic fields, you say? Magnetic fields are caused by current flow. By now, hopefully, when you hear *current* and *magnetic field* in a word association game, you come up with the answer *loop*. So let's turn our scope lead into a loop antenna by clipping the ground to the probe tip. You will see that the previous voltage signal from the outlet disappears. However, move your new sensor near the power cord of the scope you are using or some other device that is moving current. Voilà—you pick up magnetic fields with this configuration. You can often use this simple technique to determine the type of EMI you are dealing with. (And you didn't have to buy expensive sniffers and spectrum analyzers!)

Once you get over a wavelength away, the prominence of one field over the other tends to disappear and that leaves you dealing with RF, or radio frequency, interference. How do you find out if the problem is RF? Try moving the suspected interference source over a wavelength away and see whether you still have a problem.

To sum up, radiated EMI can be divided into three categories: near-field magnetic, near-field electric, and far-field or RF. The only reason to do this, though, is to identify ways to eliminate the problem. In all three cases, at some point the radiated effects have to turn into a conducted effect to disrupt your circuit. The trick is to stop that from happening.

Deal With It

Whatever the source, at some point in your career you are going to have the opportunity to exorcize the EMI ghost from your circuit. Before we get into

specifics, such as when and where to hang a juju bead,[5] there are some basic concepts that will help put these demons back in their bottle.

Break It to Prove You Can Fix It

Remember that EMI is caused by some sort of electromagnetic field, either conducted or radiated. Often this only occurs on an occasional basis. That in itself can make it hard to track down. So we will review the concept of breaking it. If you ever think you have solved a particular problem, you will need to remove the solution and see whether the problem comes back. Break it, fix it, and break it again, as we learned earlier. Due to the sneaky nature of EMI it is particularly important in this case.

Here's an example: One time I was trying to eliminate a flickering problem on a display we were using. As I worked out what was going on, I tried putting a ferrite on the wire harness. The problem went away. Thinking I had solved the problem, I instructed the production line to install ferrites on all the machines. You can probably guess what happened. Shortly after the line started up again, the flicker was back. I later discovered that the problem was caused by motor brush arcing. I just happened to put the ferrite on when the motor brushes "burned in," eliminating the noise source. Now I will always remove and reinstall the fix several times to be sure the problem returns and is eliminated consistently. The first thing I ask any engineer when he or she returns with a fix is, "Did you remove it and make sure the problem is still there?"

If you can't break it at will, you can't be sure the fix is legit.

TIMING IS EVERYTHING

Another thing I have learned is to track down the sick circuit right when it is failing. Often you might be tempted to leave it till you have time to research it. Then when you go looking, you can't find it because it's working now. You have to catch it in the act, so to speak. So when it happens, don't wait, grab your "juju kit" and go ghost hunting. Don't be surprised if something happens on the production line that you can't get to repeat in the lab. Go to the line and try to figure it out.

Amazing amounts of noise can be found on the production floor. There are usually all sorts of motors and equipment running and creating EMI on a production

[5] *Juju bead* is a term I use to refer to ferrite beads and clamps. It seemed appropriate in reference to the way ferrites seem to magically eliminate an EMI problem.

floor. A production line where I worked had a metal table that would mess up a portable CD player whenever it was within about 2 inches of the table surface. The table was grounded to a steel post holding up the ceiling. I learned that you can have upward of 50 V of noise between ground in the outlet and the steel in a building that is tied to that ground. Tying the table to the outlet ground made the problem go away. I didn't forget to try to break it by removing the fix. In fact, I did this several times just to be sure it really was the problem.

It is difficult to get an EMI problem to occur at will, so don't be afraid to go to the problem where and when it happens.

UNDER PRESSURE

Sometimes we are under pressure to develop a solution fast. To do that you might try throwing everything you've got at it at once. If you solve it, then try removing one piece at a time. EMI problems are often combinations of various things. If you try one fix at a time, you might overlook a combination of fixes that would have solved your dilemma. You might need that 0.1 μf cap on the AC line and the ferrite clamp on the data harness. As often as not, you will need more than one fix to solve the case.

BE PREPARED FOR SURPRISES

An across-the-line AC cap will do great things to filter out noise coming into your system. That's why they put them in surge suppressors. That was an absolute truth for me till a while back when I was tracking down a noise problem on a communications harness and I noticed something funny. I was observing the noise on the communication lines when I asked one of my engineers to plug the unit under test into a surge suppressor instead of directly into the wall. The noise got worse. I'm still not sure why, but we used it to improve our filtering and the reliability of the data. The moral of the story is: Don't make any assumptions. Test everything.

Not All Components Are Created Equal

What is *Xc* for a 1 μf cap and 0.01 μf cap at a frequency of 1 MHz? Let's see, $Xc = 1/(2 * 3.14 * 10\,M * C)$, so multiply, cancel the exponents, mumble, mumble, grunt, grunt. You get 0.016 Ω and 1.6 Ω, respectively. The larger cap should effectively short more noise to ground. Too bad this isn't a perfect world or that would be the case. Take a look at a regulator data book; what are the recommended capacitors? One large and one small one, right? The reason

is that the larger capacitors often do not work like smaller caps at higher frequencies. A perfect cap would, but alas, there are no perfect caps, only perfect calculations. Hint: Select a cap with a roll-off close to the frequency you are trying to clamp.

One other thing: The capacitance printed on the case is only legitimate when it's used at the operating voltage on the case of the cap. The moral of the story: You might have the right component but the wrong value—nothing a little experimentation can't solve.

Controlled Environment

Every engineer knows the importance of a controlled environment to determine the validity of a test, yet I see this concern overlooked often when I'm trying to track down an EMI problem. Maybe it is because EMI is so difficult to reproduce. There are some standard techniques for reproducing EMI in a test environment. If you have ever dealt with the European CE requirements, you might be familiar with some of them, such as EN 61000-4-4. This standard references one test that I find particularly useful: the EFTBN test. It stands for *extremely fast transient burst noise*. This is a great test for finding immunity problems with a given design.

Its history goes back to the 1960s and 1970s. Some IC-based clocks that were being developed seemed to become inaccurate during use. No one ever really located the source of the noise, but they found that if the clocks could pass this test they developed, they kept time correctly. What they had developed eventually became the EFTBN test. (It creates a similar noise profile to the showering arc test that UL used for some time before replacing it with the EFTBN test.) Remember the rusty file test from Chapter 4? This is the legit, controllable version of that.

In the same standard, you can find other test protocols, including static, line surge, and others. As you look into these standards, you will find that even the humidity of the room where the test is performed can make a difference. Fully equipping a lab to be able to perform all these tests can be very expensive, but if you do not, don't be surprised by some variation in your results. My own experience with static testing shows it to be one of the most difficult tests to repeat and get the same results. I have seen a circuit tested and seen it pass one level only to repeat the test on exactly the same board at a later date and get a different result.[6]

[6]The moral of that story was that circuits will pass static easier on more humid days.

One word of caution: Merely passing all the immunity tests is no guarantee that your design is good to go. There could still be problems that plague you. In this case you will need to develop your own internal tests that you need to pass to guarantee correct operation.

Poor Man's EMI Tests

As we discussed a moment ago, it can be very expensive to set up a completely controlled test lab. Renting time at one isn't cheap, either. So, what do you do if you don't have much of a budget? Throw your arms up and forget about it? Though that is certainly appealing (especially when you are really stumped on a particular problem), it usually isn't an option.

There is a rule that crops up time and time again in every discipline that I have studied. It is the 85/15 rule (you might have heard 80/20 or 90/10). What it means is that it takes 15% of the effort to get 85% of what you need and 85% of the effort to get the last 15%. This is true in the world of EMI as well. Even if you do not have a perfectly controlled environment, you can still learn something about EMI. What you will not get is a definite pass or fail conclusion.

I have already mentioned the rusty file test as a cheap and dirty version of an EFT machine, but it's not as controlled or even anywhere close to being as safe. It is a poor man's version of the showering arc test. (The showering arc test was used by UL for some time before it was replaced by the EFT test.) I take no responsibility for injury caused by being so poor that you have to use the rusty file test, and I do not recommend it. Personally I think you should get your company to cough up the money for an EFT machine. You will have to spend a few grand, but you can get a lot from that without all the expensive shielding room and environmental control equipment. Besides, I will sleep better at night if I don't have to worry about engineers rubbing wires on rusty files.

I have heard of cheap and dirty static tests using piezo igniters out of barbecue grills; they pump out 15 to 20 kV in a static jolt. You can get about 5 to 10 kV with a nice pair of Lycra shorts on a dry day. (Beware, though—you might get some funny looks from coworkers if they see you shuffling around in your biker shorts and stocking feet carrying a PCB to test.)

Again, you can purchase a static gun for a lot less than you can get the whole humidity-controlled room with grounded floor, and you'll get 80% of the controllability that you need.

Line surges can be created by switching AC motors on and off with a simple switch. An AC fan from Wal-Mart is a common source of this type of noise.

Again, you won't be able to control the level, but you will get an idea of whether or not your design can handle EMI at all.

In general, you should do what you can to check your design. If possible, spend some money for some equipment to test, but you don't have to dive in whole hog to get some benefit out of EMI testing. This way, you can do most of the improvements at your lab, saving time and money when you take it to a certified testing lab.

I Dream of Juju

Experience is of great value in the battle against EMI, but you don't have to learn all the courses the hard way. You can learn from others' mistakes. Read what you can on the subject, but beware: There are many different opinions on this topic. Don't take what you find as gospel in your particular situation.

By its nature and complexity, EMI can be a bear to handle. You will find that some solutions won't work as well for you as they do for other people you read about. The best way to deal with this is to document your reasons and conclusions for a given fix you have found, refer to it, and update it often. Make yourself a "Juju journal." (Yeah, sounds a lot like keeping a lab book, doesn't it?) You will find after a while that there are some solutions that work particularly well for your product. Armed with this information, you will solve these problems faster and more cheaply than before. You will even begin anticipating ways to avoid them after a while. I have even woken up in the middle of the night with the solution in mind. Don't overdo it, though; you don't want all your dreams to be of Juju beads and PCBs.

It's in the Air

If you are trying to stop EMI out in the air, your most likely solution will involve some type of shielding, which means putting your design in a conductive box. If it is RF, you will need to keep the holes in the box smaller than the wavelength of the signal you don't want.

If it is near field, there are some variations on the box. Sometimes all you need is a grounded plate between the circuit you are trying to protect and the source of the noise. For magnetic fields or current effects, ferrous shielding works well. For voltage or capacitive effects, something simply conductive will work. Whatever your approach, if you try to stop it in the air, it will involve some type of shielding and will very much be a trial-and-error process. It is also the most costly solution. For this reason, I tend to treat shielding as a last resort. I go to the wire first.

It's in the Wire!

At the end of the day, all EMI is conducted. EMI can't disrupt anything until it is conducted. Even when you are dealing with near-field and RF disturbances, when it is all said and done, unless it disrupts a signal on your board, it doesn't matter. That alone makes learning how to deal with conducted EMI important. It also means that the board and circuit design itself can affect EMI tremendously.

Here are some rules of thumb in PCB and circuit design that you can use to stop EMI in the wire.

Low Current (Power) Signals Are Disrupted Easily

Signal-to-noise ratio is based on power, both voltage and current. Mostly we work in a world where we keep voltage the same and current is allowed to vary. That combined with a need to conserve power often leads to some very low-current signals. The problem is, if the signal is low in power, the corollary is that it won't take much power to disrupt it.

For example, you can stick your hand in a stream from a 49-cent squirt gun and easily deflect the water, disrupting the signal. Try doing that with a fire hose and you might lose your hand.

In most cases, radiated signals don't have much power behind them once they are absorbed into your circuit. That makes it easy to combat them in one simple way: Make the circuit under distress use more current and thus more power—turn it into the fire hose so it can't be easily disrupted.

Take a sensor with a 1 meg pull-up at the end of a 4-foot wire. Change the pull-up to 10K and watch what happens. This is one reason that the old 4/10 mA current loops are so darn robust. They are hard to disrupt.

If you really can't spare the extra current, you will need a component that has a low impedance at the frequency you are trying to suppress and a high impedance at the lower frequency at which your signal is operating. They have those; they are called capacitors. Putting one of these on back at the input of the device in question will create a load at a specific frequency, making it harder for the unwanted signal to disrupt the wanted signal.

Find the Antenna and Break It

Increasing power to a circuit works great unless the signal causing you fits is at the same frequency as the signal you need to read. When this is the case, you need to consider antennas.

In a very real sense in the world of electronics, everything is an antenna. The only question is, how good an antenna is it? But first, what *is* an antenna?

An antenna is a device that turns a radiated field into a conducted signal. There are two basic types: the dipole, a ground and a length of wire, and the loop—you guessed it, a loop of wire. Earlier we learned how to turn a scope lead into both types of antennas to discover some of the EMI in the world. The loop is particularly good at picking up magnetic effects, whereas the dipole does well with capacitive effects. At RF levels, there are all sorts of equations and loading formulas that are more in depth than the scope of this text. Suffice it to say that RF can be picked up with both antenna types.

The trick is identifying antennas in your design. Once you find them, you can figure out what to do with them.

Sometimes you might identify an unknown antenna in your circuit when you are checking for conducted effects. You might unplug a long wire, for example, and discover that the problem goes away. I have had this exact thing happen more than once where I unhooked some contacts that were getting a static discharge, only to still have a problem. The problem only went away when the wires that routed out to these contacts were unplugged. I had removed the antenna.

Dipole antennas tend to be wire harnesses that plug into the design. One way to hamper these antennas at higher frequencies is to put a ferrite bead on them. Now you know why those little bumps are on so many wires these days.

Loop antennas are often found right on the PCB. The higher the frequency, the smaller the loop needed to have a problem. In general, the smaller these loops, the better your design. An easy way to improve this, if you have money to spend, is to go to a four-layer PCB with a ground plane and *Vcc* plane on the center two layers. That way you always have the smallest loop area. If you don't have the bucks to spend on a four-layer board, it will take some practice and patience to learn how to do the same thing with a single- or double-layer PCB. I highly recommend a class on this topic for your PCB designers if this is the case. There are many available lecturers on the subject.

As a general rule, good radiators are good receivers. This being said, you can turn your circuit on and, using the scope probes, find hot spots on your PCB or wire harnesses and get an idea of where the trouble is. If you need to be more precise, you might want to invest in some near-field and sniffer probes for your equipment. Find the antennas in your circuit and break them (make them bad antennas) to stop EMI.

In Conclusion

There is no simple approach to dealing with EMI, and experience rules in this arena, so don't be afraid to get your hands dirty trying to figure this out. Also, there are many texts out there on this topic and this discussion is by no means comprehensive, but I will warn you that not everyone agrees on the same approach. You will need to find out what works for you and your product and go with that.

One final note: The things you do to keep EMI out will also keep it in when you are trying to pass those emissions standards that seem to get tougher and tougher, with no end in sight. Make your circuits more difficult to disrupt, ferret out those unknown antennas and break them, and when all else fails, shield your circuits.

Use the following Thumb Rules to help you exorcize the ghost that's in the machine.

Thumb Rules

- EMI comes in two flavors: conducted and radiated.
- Radiated effects can be divided into near field and RF.
- Near-field effects can be magnetic or electric.
- Identifying the type of EMI you are dealing with can help you develop a solution.
- Start with unplugging and unhooking whatever you can.
- Is the fix repeatable? Can you break it by removing the fix?
- Chase down the problem where and when it is happening.
- Remember, components aren't perfect.
- Keep a log of solutions.
- Low-current signals are disrupted easily.
- Find the antenna and shut it down!
- Load the dipole antennas to stop electric fields.
- Minimize loop area on the PCB to stop magnetic fields.
- Good radiators are good receivers.
- When all else fails, shield it.

CODE JUNKIES BEWARE

Our world relies more and more on software. In saying this, I include firmware, which is really software that you simply don't change as often. It is in everything. Even good old analog circuits are evaluated by software in most cases. This is a good thing because of the flexibility that it has created and the new features that are available (my home stereo wouldn't be the same without DSP!), but it comes at a price. The world of buggy software we live in today is that price.

Bug-Free Software Might Be Impossible

If we are talking 20 lines of code, we can make it bug free, but what about a million lines? Or even 1000? The more code there is, the harder it becomes to make it bug free. I have no proof as Einstein did, but I think it is akin to the law of relativity—the closer you get to the speed of light, the harder it is to get there, basically making it impossible. Likewise, the more code you get, the harder it is to make bug free.

Whether your code is 50% bug free or 99% bug free depends primarily on one thing: how much time you have tested it. The more features and complexity in the code, the more time is required. At some point you have to figure out a balance between a level of bugs you can live with and when you need to ship the product. Since we as consumers now demand everything at the lowest possible price, we have created a world of upgradeability. You can buy my possibly buggy program now and upgrade it later. This even happens in everyday devices, not just computers. I have upgraded my PDA several times, and I just found out there is a new version of OS for my PSP. I have even upgraded my GPS unit a couple of times and I couldn't tell you how many times my iPod has been updated.

So, if your code is gargantuan and you want really bug-free stuff, your cost will be high and it will take lots of time. Space Shuttle code is up there in the bug-free realm, and it is possibly the most expensive code per line ever written.

This is why big software companies that start with letters like *M* sell you code that you never truly own and aren't responsible for it malfunctioning. To guarantee it would simply be so expensive that no one would ever buy it. Software never can be truly perfect, but it can be good enough. "Good enough" is completely subjective, however, and it is up to you and your company to determine what level that is. Here are some ways to troubleshoot your code and help you determine whether it is good enough to ship.

Testing, Testing, and More Testing

Good code takes a lot of testing, if you hadn't gotten that idea already. I particularly like human testing, where the person who's going to use it is involved. We humans always seem to discover ways to break stuff that you simply didn't think of when you designed it.

The problem with human testers, though, is getting them to remember what they did when it broke. Memory can be a fickle thing, and when you are drudging through a test, exactly what you did when the unit malfunctioned is likely a poor recording. In one place where I worked, we put cameras in the test lab to watch the human testers so that we could back up the tape and look at what happened. It saved us from chasing down more than one dead end.

Repeat the Problem

Like most difficult-to-trace problems, the ones that are hard to repeat are the hardest to find. With software, it is not unusual to have a certain set of conditions required for the bug to manifest—certain key-press combinations, or maybe timing. If you are chasing down a bug and you just happen to make it repeat, stop, rewind your brain about 30 seconds, and see if you can do it again. Keep trying slight variations on whatever it was that made the bug show up until you get it to happen again, and then try to repeat it one more time. Keep trying till you can get it to happen whenever you like. If you can get it to happen on cue, you will be able to track it down much more easily.

Set Up Tracers

In code it is possible to set up tracing registers that can keep track of key information that will help you figure out what went wrong. This can take up some extra time in development, but it will pay huge dividends in the debugging process.

One time we had a problem with a control panel resetting at apparently random intervals. We checked the stack by creating a register that kept track of how deep the stack would grow. As we watched it, the stack would get so big it would overwrite other areas of the code and it would go into "la la land" till a watchdog timeout reset it.

Often you can use an available display to show this information. However, there are times when you might want the information faster than the display can update, or maybe the display can't show you what you want to look at. In this case you should set up a D/A—some type of circuit or signal that can take

any register in your micro and turn it into an analog signal that you can hook a scope up to.

You have to debug this and gain trust in it before you use it. Do so by loading any number into a known register and look at the scope and see whether it is what you expect. Once it is working well, you can use it to do the same type of root-cause analysis as the hardware guys. You methodically plug each number into it at various stages of calculation and work your way back from the offending output till you find the cause of it all.

This method can be used with simple RC circuits, serial D/A, or any myriad of options. Some chips even have some tracers built right into them. The point is to follow the same root-cause analysis as previously discussed, but in this case you need an idea of what is happening inside the chip at any given point in its processing.

Code Reviews

One way to debug code is with a review process: Put your code up onscreen in front of several engineers, and have one engineer (specifically not the one who wrote it) walk you through the code step by step. If you can overcome the natural tendency to nod off in this type of meeting, it can be quite effective. I suggest using it for specific cases and keeping the review time short, since focus on understanding what the code is doing is paramount to making this work.

Break It Again

Just as we already learned, this is a great way to make sure that you have fixed the problem in software as well as hardware. If you can break it and fix it at will, chances are that you have found the bug. This is much easier with the advent of Flash chips. In the old days of OTP manufacturing and EPROM prototyping chips, you had to wait forever (nearly 20 minutes, can you believe it?) for them to erase under UV.

Hunting Bugs

Even though I'm a diehard analog guy at heart, last time I looked software wasn't going away anytime soon. So we do have to live with the fact that my DVD player takes longer to boot up and read a disk than my TV took to warm up its tubes 30 years ago.

The fact is, code has become a way of life. We are even teaching our children how to handle the convoluted and twisted thinking you need to write code.

Just take a look at the video games they are playing! I think I need an upgrade to my noodle just to play them, and I dumped more quarters in the arcades than most of my peers did years ago.

Enough reminiscing. Software is here to stay, and unless the Internet gains consciousness sometime soon and can debug itself, it is up to us, so good luck on the bug hunt!

Thumb Rules

- Test a lot, record information somehow, and don't rely on human memory.
- Rewind your brain 30 seconds and try to repeat the problem.
- Set up tracers; use what the chip has, and if not there, build in your own.
- Use code reviews to explain and review your thinking.
- Repetitively break it and fix it to prove that you have found the bug.

CHAPTER 7

Touchy-Feely Stuff

This is the touchy-feely part of the book. Before you say "Ick!" and chuck it as far away as you can, please read on. Most "average" people find the people who gravitate to the world of electrical engineering a strange lot. If it weren't true, *Dilbert* simply wouldn't be funny. From the point of view of the EE, the rest of the world often doesn't seem to "get it." If you want to be the most successful engineer you can, there is some touchy-feely stuff you ought to chalk up on your list of acquired skills. Yes, it is extremely likely that these are going to be acquired skills; the engineer who comes by these capabilities naturally is a rare breed.[1]

PEOPLE SKILLS

One difficulty engineers often have in dealing with people is the fact that interactions between us human beings can't be described by slick mathematical formulae like the various circuits they are working with. I personally think this is why you often see engineering groups managed by nonengineering majors.[2] So, what should you do? One thing I have found is that, though there is no

[1] I might venture that these are the people like Steve Jobs, Bill Gates, and other famous and now very rich self-avowed techno-nerds. Maybe if you can grasp these skills that type of success will be in your future. Trust me, it is much easier for you to develop socially than it is to teach those other types how to pick apart a circuit!

[2] If you dig deeper, you will find that these types of managers can somehow interpret typical "sparky speak" into language that is less analytical and somehow more socially skilled types can handle.

perfect equation to describe people, there are some categories into which you can sort people to help you understand how to interact with them.

In any business organization there are levels of hierarchy—there is no round table. Someone sits at the head and it goes down from there. There is always a pecking order, even if it isn't on the org chart. Let's sort the personality categories into various levels of interaction, since that will definitely affect how you should react. We might as well start at the top.

Note that I am using masculine pronouns in these people descriptions for convenience only. Of course, all these people can be either male or female. Maybe someday we'll invent some effective gender-neutral pronouns. Until then, please feel free to use the pronoun that offends you the least or makes you laugh the most!

Those Over You

This means your boss, the person you report to, and the person who takes responsibility for what you do. Of course, that is in a perfect world.[3] First, some general rules:

- *Avoid talking smack about your boss.* Even if he deserves it, constant griping and complaining will usually hurt you more than him.
- *Maintain integrity.* Sometimes lying and deception can get you ahead in the short run, but in virtually every case it will come back to haunt you.
- *Help your boss succeed.* This can be hard sometimes, especially if your boss never gives you credit, but even if that is the case, be a great employee. Someone will notice.

The following sections contain descriptions of some boss types.

THE *DILBERT* BOSS

This is the clueless boss. He has no idea what you do, and he is more concerned with his position than with the success of the company. He is more than willing to sacrifice one of his employees to make himself look good. This is the type to take credit for everything you do right and blame you for everything that goes wrong. First, do the best job you can. Your boss's own self-interest

[3] I am well aware that there are plenty of boss types who will take all the credit when you do good and lay all the blame at your feet when you screw up. I truly hope you are never saddled with such a boss, but read on for some rules that will help if you are.

will keep you around if you are a valuable employee. Second, look for opportunities where others in management can see your skills. This will counter the fact that your boss tries to hide you away. Transfer out of this group if you can, since it will be difficult to get far with this boss.[4]

NEGOTIATOR BOSS

This is the salesman type, the supreme negotiator. He will always set the goal beyond any reasonable point, figuring that somehow this will encourage you to go further than you think you can. First, don't be discouraged by these requests. After that, you have two approaches you can take. Be a negotiator yourself—overestimate the time and money it will take to get the job done so that you have room to negotiate (like Scotty does for Captain Kirk on *Star Trek*). The other option is to say what you can do and stick by your guns. Don't underestimate with the negotiator, though—he will be disappointed when you don't meet the goal you said you would. The negotiator is not necessarily a bad boss to have. You could do much worse. "Better to aim for the sun and miss than aim for a cow pie and hit it," is the creed of this boss.

THE "YES MAN" BOSS

The "yes man" is the submissive boss. He tells his boss anything he wants to hear and will often not defend his employees. It is not unusual for this boss to commit you to impossible deadlines and tasks. Don't make the mistake of being a yes man to a yes man, though—that is a disastrous combination. Let this type of boss know what it is really going to take to get the job done. If you have a strong personality, you can help this boss by standing up for him if he does say what it's going to take to get the job done to *his* boss. Generally this boss will give you the credit for both your successes and failures.

THE MICRO MANAGER

The micro manager tries to manage every detail. Try to handle his status report requests and required updates as quickly as possible so that you can get back to business. He might even be so obstinate as to be upset when you try to make a decision for yourself.

I think the best way to deal with this type is to simply make sure you get those reports and updates in on time. Try to be so reliable that this boss will gain

[4] This is the "dud" that you will learn about later who was somehow promoted to management. Yes, that happens, and if it is prevalent in your company, it's best that you start looking elsewhere for employment.

trust in you. Often you can talk to this boss easily (there will certainly be enough meetings with this guy). Talk to him about your priorities often, and stay in sync with his goals for the department. As long as he doesn't carry it to extremes, you are better off with this guy than the *Dilbert* boss. So don't feel too bad for yourself.

THE MACRO MANAGER

The opposite of the micro manager, the macro manager is the boss who is never there when you really need some help. He is hard to get hold of and often difficult to talk to. This leaves you making a lot of decisions that you might not feel comfortable with. You might even be criticized for decisions you've made after you asked repetitively for some feedback on that particular issue without response.

The best thing to do in this situation is to take advantage of the opportunity to learn to make decisions on your own. You might screw up, but that is a risk you take in any decision situation, so don't be afraid of making a mistake. If your boss does question your reasoning, try to explain your decision process. Remember, he wasn't there for all the things that led to your choice. Don't assume that he has the background on the issue that you do. The best thing about this boss is the opportunity you will have to shine. You will be given plenty of rope; try not to hang yourself!

THE PERFECT BOSS

The best boss gives you some credit while buffering you against mistakes, giving you a chance to learn and grow. If you have this type of boss, do your best to succeed and you will! You should hang on for the ride. Often he will give you plenty of leeway to succeed. He will recognize that his success depends on yours, and he will help you succeed. Don't be upset if this boss gets some credit for something you did. If he is a good boss, he created the environment that allowed you to be a success and deserves a nod for that. Often, as this boss succeeds, you will as well because he will bring you along with him.

YOUR BOSS'S BOSS

You might not get a lot of interaction with your boss's boss, but take care when you do. This is the most visible you will get as an employee. Try not to be too nervous. I remember one time I was dealing with the CEO of the first company I worked for. Our production line was shut down because of an electronic power problem. I was a lowly part-time student tech in the QC department. I had just figured out the problem when he came to the line to see what was

up. I was shaking in my shoes as I showed him the cause of failure. He didn't believe me at first, so I showed him a broken one, fixed it, then broke it again. He was satisfied, and production started back up. It only took two or three more of those situations and the CEO knew my name. If I had panicked in that position, no matter how right I had been, the results for me would have been a lot worse.

THOSE OVER YOU SUMMARY

A point to consider with these categories is that it is possible to find variations of these types. After all, as we said originally, this people stuff isn't an exact science. If your boss is a blend of these types, you will probably have to blend your response as well. If it helps, make up your own boss type; figure out his or her attributes and what seems to make him tick.[5] Use what you figure out to guide your choices.

Those at Your Level

Ah … your coworkers, your fellow peons, and sometimes your enemies. This level of interaction with your network of peers is the best place to create future opportunities. The following subsections describe some peer types.

THE SNEAK

Watch out for the sneak. He is always trying to see what he can get away with. He will only work hard when the boss is watching. Don't get caught in any of his schemes to take advantage of the company. That usually turns out badly and gets you branded as a sneak as well.

THE POWER MONGER

A true political figure at work, for the power monger it is very important to build power and reputation. What is sad is that he might try to make you look bad to make himself look good. Try not to give him any ammunition that he can use to prove how badly you are doing, thus making him look better. You can make alliances with this guy pretty reliably, but it will be an "I scratch your back, you scratch mine" type of relationship. If you make deals with this person, you will need to hold up your end of the bargain, since you will be relying on his self-interest to hold up his end.

---

[5] Take caution to not expect the same behavior every time. Remember, people aren't as predictable as a circuit. Even so, this can be a very effective exercise. It will help your career more often than not.

THE BADGER

The badger will tend to respond emotionally to situations. If she feels she is being attacked, she will likely get defensive and angry like a badger when cornered. The best thing to do is back down and give her a chance to calm down. If you can help this person get past the emotion (or just wait it out), you can usually reason with her. It is not unusual that the badger is also a workaholic. Maybe that is why badgers are so ornery.

THE AVERAGE JOE (OR JANE)

Companies are filled with average Joes. These people do a decent job, nothing stellar, but are fairly reliable. I believe that if it weren't for average Joes, companies could never be formed and kept together. These people like the security of someone else making the tough decisions. They will often ask you what they should do. Average Joes like to look to a leader. If you can gain their respect, others will notice and it could lead to a promotion.

THE SHOOTING STAR

These are the guys (or gals) who get it. They work hard but don't make themselves into badgers. They are reliable and often correct in their decisions. True shooting stars possess integrity and a desire for the company to succeed. They often get promoted as these skills are recognized. This is a good friend to have in a company, but hopefully after reading this book *you* will be the shooting star that everyone else wants as a compatriot!

Often the shooting star is a leader and a true mentor; even if the organization chart doesn't show it, you should listen to the star's advice whenever you can.

THOSE AT YOUR LEVEL SUMMARY

One of the most important things to have at this level is respect, for yourself and for the others you deal with. You gain respect for yourself by following through with what you say you will do. Stick to your word. If you make a mistake, say so, correct it, and move on. Give others a chance to build respect at the same time as you. This mutual respect is a way to build a network of contacts that is synergistic in nature. Here is where you and your colleagues can help each other out, do each other favors, and be more successful than you would be on your own.

Those Under You

You might be looking for a chance to lead or have had it forced on you. Either way, you ended up with some subordinates who answer to you. This is

commonly the hardest adjustment for the true engineer type. As these people below him on the org chart interact, he or she will be baffled at the behaviors and personality traits that come out. Here are a few buckets to sort them into.

THE SMART SLACKER

Smart slackers are usually pretty smart and can get a job done more quickly than most others in the group. For this reason they might get some free time when others don't. But they don't go looking for any more work—they goof off or spend the time surfing the 'Net or other such things. Usually they are quick enough on the keyboard to get back to looking busy when you walk by. Keep their plates loaded to the brim. If their slacking becomes a big problem, you might need to call them in and discuss it.

THE PRAISE DEPRIVED

Praise-deprived employees often need daily feedback on how they are doing. They are looking for positive reinforcement and need a little praise. Be sure to let them know when they are doing a good job. Don't be afraid to be constructive if they make a mistake or should try a different approach. They will usually let you know if they are done with a task and need more to do.

Sometimes as a boss, you will wish they would just leave you alone, since they can seem a little needy. If they are valuable employees, spend a few minutes with them as needed. They will be very loyal for that little time you spend. If they aren't so good, ignore them and they will find a job elsewhere, solving the problem for you.

THE DUD

The dud is the person who doesn't bring a lot to the table. You have to put more work into him than you are getting out of him. That said, I am a firm believer that people can change and improve. I prefer to give them a chance, but be firm in laying out the expectations. Let him know what is needed from him to keep him employed.

This, however, is not a situation that you can keep dealing with forever without draining resources from the company. If he doesn't change, this is the person that you have to make a hard choice with, the one you have to let go. Don't run your group with a drain on resources indefinitely. It will hurt all of you in the long run.

THE AVERAGE JOE (OR JANE)

This is the same guy we talked about earlier. Be a leader for him, show him how to excel, and you just might turn him into a shooting star.

THE SHOOTING STAR

The shooting star is the same kind of person we already discussed. Most important: The more of these you have in your group, the better you will perform! Don't be afraid of giving him or her credit, and don't try to suppress any one of them into being your peons. It will backfire on you. Share the credit and hook your wagon to these people; they will get you to the finish line!

FINALLY

Can a truly effective manager get an average Joe or Jane to become a shooting star? Or make a dud into something more? I think so, and I believe it is the mark of a good manager to do just that. Anyone can yell and intimidate people into doing what they want. The manager who persuades and edifies is much rarer and also more valuable. His or her team will be more efficient, have less turnover, and just get more stuff done. It doesn't mean you should be an old softy. You might need to be firm at times, but if you truly care about your employees, it will show and make a difference.

Administrative Assistants

Every organization has an underground method of communication. In most companies it flows through the assistants. Building a good rapport with the secretaries and assistants is a good idea. It will allow you to tap into a whole other communication structure. If they think well of you, you will have a better reputation with those above you. Help the assistants whenever you can, and treat them with respect. A lot of unsung greatness lies with the assistants in an organization. This applies to your assistant if you have one. Don't ever degrade them; it will come back to bite you. If they respect you, it will proliferate through the network and help you. If they don't respect you, that will travel the network and hurt you. This doesn't mean you just let them goof off all day. As individuals they will fit into the categories we've described and can be dealt with similarly, with respect.

BECOMING AN EXTROVERTED INTROVERT

It seems to me that, generally speaking, the personality types that do well in engineering seem to be naturally shy. I would have to say that electrical engineers are probably the most introverted of the bunch. I was once asked, "How do you tell whether you are talking to an extroverted engineer?" The

Thumb Rules

- Work for the perfect boss when you can; work with what you get when you can't.
- Gain the respect of the average Joes.
- Hook up with the shooting star.
- Be a shooting star yourself.
- Blend your response to blended personality types.
- Give the dud a chance; let him go if he doesn't step up.
- Make the average Joe into a shooting star.
- Treat the administrative assistants with respect.

answer: "He is looking at your shoes, not his own." It's funny because it's true. It is also true that the EE can benefit by overcoming this tendency. Here are a few ways to do just that.

It All Depends on Your Point of View

A wise man once said (and I'm paraphrasing), "You will find that right or wrong often depends on your point of view."[6] Given that, I will try giving you an idea of the way things are seen from the most common sides of the fence. For this discussion we will call the engineer the peon and the manager the pointy hair.[7]

THE PEON POINT OF VIEW

The decisions and directions of management are often as undecipherable to the typical engineering peon as ancient Egyptian hieroglyphics are to the average person. Here are some insights into the thoughts that go through a typical EE's head when dealing with a pointy hair: "Why in the world is this the most important thing now when just yesterday it was the last thing on the list?" Or maybe, "Why can't you understand things like the word *impossible*?"

In my early years as a peon I coined the phrase, "Management is an unnecessary evil." It accurately summed up my thoughts on the topic. If your manager couldn't help you with fixing that circuit that wouldn't work right or the code that just didn't execute the way it should, what good was he? I mean, sure,

[6] Obi Wan Kenobi said this; some great life lessons can be learned from *Star Wars*!

[7] Yeah, I keep lifting phraseology from *Dilbert*. What can I say! Scott Adams struck a chord that rang true throughout the corporate universe!

he could keep buggin' me all the time about status reports and the like, but couldn't I manage my own time?

Even if you find engineers with a manager that they like and think is very helpful, they are still at a loss to understand management decisions. This is often due to a lack of background on the decision process. Good managers will often help this situation with some explanation as to the way they came to the decision. Engineers, though usually a little underdeveloped in the social skills area, still understand numbers and reasons. It helps them to know *why*.

There is a natural angst in the role of the engineer vs. the manager. After all, he is the peon in the relationship. At the end of the day the manager is his boss, not the other way around. Remember, engineers spend their whole lives asking themselves, "Why this?" and "Why that?" It is what they are trained to do; it makes them good engineers. Help them answer that question if you are a manager!

THE POINTY-HAIR POINT OF VIEW

First, understand the first and foremost goal of a manager: It is to make the business successful. It's either that or to make the department he is managing look good, which coincides with the first unless it is sacrificed for the second. (This can happen with bad managers. Hopefully their bosses will notice and correct that before it is too late.)

The good manager wants a successful company; how do you do that? It is pretty simple really; you make more money than you spend.

Where the engineer is more focused on accomplishing the task at hand, the pointy hair worries about getting it done on time and on budget. This often puts the peon and the pointy hair on opposite sides of the fence. It is difficult for a manager to understand that unknown things can come up that mess up the estimated schedule the peon gave him. Here's an actual quote from a manager: "We need to figure out a way to predict unknown problems from happening and avoid them." He was completely serious.[8] To him, that is how to get from point A to point B. To the engineer who is trained to think logically, this phrase will cause his brain to strip a few gears, leaving him generally speechless and unable to respond.

It is not a bad thing to think so far out of the box.[9] If the engineer can shift his head back into gear after such a phrase and look at it as a problem to solve,

[8] I was personally flabbergasted at the time; this was before I developed my personal understanding of the pointy-hair point of view.

[9] I've seen pointy hairs so far out of the box that I wasn't sure there was even a box around!

you will be surprised at what you think up. It is logically true that you can't predict things you don't know. However, you might come up with a way to find out some things you didn't know before, and avoid those. Which is what that "pointy speak" really means.

When two engineers start talking, you will often see pointy-hair eyes glaze over as if you were speaking a language they don't understand, which you are. To keep them interested, use words like *schedule* and *budget* a lot. Managers like to speak in absolutes, as in "This will be done in such and such time and cost so much." Engineers like to have some fudge factor. They have seen too many failed lab experiments to believe it will always go right the first time.

In my experience, if you tell the pointy hair it will cost between 10 and 15 bucks, the only price he hears is 10 bucks. This being the case, if you aren't sure you can get to the low price, you'd better not say it, no matter how often he tells you not to sandbag your numbers. If you have some confidence, though, go for it—it is also the mark of a good engineer to get to the committed price and schedule, even if it takes some extra effort. Take caution, however—you don't want to sandbag a number so high that you never build anything because it is always too expensive. Remember, the goal of a business is to make money, and you can't do that unless you make stuff and sell it.[10]

TALK IT OUT

If the engineer makes an effort to lay off the acronyms and the manager tries to explain some of the reasons behind his decisions, it will do wonders for your mutual understanding. The most important thing you need is a desire to understand each other. We'll get into the skills a little later.

Visualization

A few years ago, as I watched an interview with Michael Jordan, I realized that we have something in common. No, it is not a 40-inch vertical leap or the ability to dunk the ball. I realized that for years I had been using a method for success that Miracle Mike also used, a technique called *visualization*.

Everyone who works for a living experiences difficult and stressful situations. It might be dealing with an irate boss, a lazy employee, or a fellow manager who just doesn't seem sane. Have you ever left a difficult situation in which you were trying to argue your case when you suddenly thought, "I should have

[10] Unless you are doing government work, and that is a whole other philosophy!

said blah, blah, blah or yada, yada, yada?" You might be saying to yourself, "Hindsight is 20/20," but what I am about to tell you is how to turn that hindsight into foresight.

I remember one of the first conversations I ever had with a CEO. I was a lowly engineering student; he was the boss of a $700 million company. He hit me with a couple of questions that I was not prepared to answer. I still remember how my mind drew a total blank. Afterward, as I thought about it, I knew exactly what I should have said. I decided that I would not go into such a situation unprepared again. But how do you prepare for something like that? This is what I did.

I started to imagine myself in the situation beforehand. I would imagine how the conversation might go. He would say "this" and I would respond with "that." In my imaginary situation I would try out several different approaches and then imagine a response. I would visualize the person understanding my point and a resolution to the case at hand that I desired. I found that when I did this, the real conversation, when it occurred, followed my imaginary one so closely that I always knew what to say. And better yet, I usually got what I wanted out of it.

You might think I am full of it, but I have used this technique to visualize getting raises and promotions, and I can honestly say that I got what I asked for in nearly every situation. It actually amazes me when I look back at it. I was promoted into engineering positions when I was still a student. Later I worked with several people, including a former boss, as an equal or superior. I could hardly believe this happened to a naturally shy person from a hick town in Utah, a person who doesn't like confrontation.

There are no set rules for how to do visualization other than the more often you do it, the more successful you will be with the technique. If you imagine the ball going in 1,000 times, the next time you have to shoot that clutch shot, it will go in. It works for Mike, and it works for me. Give it a try.

Affirmations

One of my favorite *Saturday Night Live* skits is the one where Stuart Smalley says, "I'm good enough, I'm smart enough, and doggone it, people like me!" He mocks a technique similar to that of visualization. It is called *affirmation*.

If you get into quantum mechanics, there is a rule called the *Heisenberg uncertainty principle*. It was developed to understand some interesting experimental results in which quantum particles (everyday light being one of these particles) act like a wave in one experiment and like a particle in another. The problem is,

they can't be both all the time; the behaviors are mutually exclusive. Anyway, a general conclusion of this principle says that when you measure something at the quantum level, the very act of observation affects the outcome of the measurement. You, the observer, basically get what you are looking for.

Please bear with me for a moment while I digress into very unengineering-like metaphysical rumination. If you get what you look for, can you affect the outcome by looking for what you want? This is what affirmations basically say you can do. Affirmations are a lot like the visualization technique we discussed but taken to the next level. You not only imagine what you are going to say or do in a given situation, you imagine the outcome you want.

I know it sounds hokey, and I admit that it is not a perfect process, but I also believe it works. Take any goal you want to achieve and write it down 30 times every day, such as "I will get a book published," or "I will get a raise," for example. Give it six months and see what happens. My experience is that it does work; you're reading this book, aren't you? Guess how I started that ball rolling.[11]

One thing that definitely happens when you use affirmations is that your mind spends considerable time pondering what you are looking for. This, I believe, leads to recognizing opportunities when they come your way so that you act on them. Several years ago I had on my long-term goal list a desire to publish a book; it was a goal I affirmed regularly. I thought about it a lot. Then, while reading an electronics magazine, I saw an ad for writers. I sent in a reply and they asked for a copy of my work. I sat down and wrote my first column. It was a success and I began writing. One opportunity led to another and here I am achieving the goal I had set out to do. Imagine, however, if I hadn't had this on my mind when I saw that first ad? Would all of this have happened? I don't think so.

You get what you look for, so control your destiny. Say to yourself, "I'm good enough, I'm smart enough, *Insert your desire here*, and doggone it, people like me!" Works for Stewart, works for me, and it will work for you, too.

Breaking Out of Your Shell

These techniques work very well in helping the naturally shy person to break out of his shell. If you can overcome the natural shyness so common to engineers

[11] It is no coincidence, in my opinion, that the techniques of visualization and affirmation mirror that of faith and prayer so closely. I think they are principles that simply work.

and learn a bit about the people around you, it will give you career opportunities you might never get otherwise.

The hardest part of breaking out of your shell is taking that first step toward doing it. You have to make the first step. After that each one becomes easier. For example, you need to make a phone call that you really don't want to make. You have already spent plenty of time visualizing it; you have thought through all the possible things that might happen. Now you are stuck—you simply don't want to make the call. It is not uncommon to feel apprehension at this point. Don't give up hope, though; there is a way through it.

First, clear your head and stop thinking about what is going to happen; concentrate on one thing—picking up the phone. Once the phone is to your ear, worry only about dialing the number, nothing else. After someone answers, worry only about initiating the conversation. Once it starts, the preparation you did with visualization will kick in and from there on things will get progressively easier.

Repeat

Though it will get easier each time you go into a specific situation, these are not skills you can learn once and then forget about. They require repetition, a *lot* of repetition, not unlike learning to play an instrument or speak a different language. The more you use them, the better you will become at them. Find out the way these skills work best for you and practice them.

I still encounter situations today where I use these skills that are more than 20 years old for me. They still work, and I keep finding new ways to apply them. And yes, I still get nervous when it is time to talk to the CEO, but now it goes much more smoothly.

Thumb Rules

- To the engineer, many management decisions don't make sense unless they know the why behind the what.
- Managers have difficulty understanding techno-speak.
- Talk it out till you understand each other.
- Visualize the situation, what is going to happen, and what you will say.
- Write your goals down 30 times a day till you achieve them.
- Break out of your shell one step at a time.
- Practice makes perfect; keep working on these skills forever more.

COMMUNICATION SKILLS

Engineers are notorious for their poor communication skills. I was once asked why it is that engineers who deal in logic that is either true or false have such a hard time answering yes or no to a simple question.

It is something that I myself am plagued with. Given a typical question, for example, "Will such and such project take long?" my answer usually begins with, "It depends …" If I'm not careful from that point on, it can quickly lead to my listener's eyes glazing over.

When you are a better communicator, you will be more successful. Simply put, everything we do in the world today requires communication. It is somewhat ironic that things that enable communication to be better (like the Internet, for example) are designed by engineers who could often use a course on the subject themselves. So here are some pointers.

Verbal

The brunt of day-to-day communication is verbal. It is also the hardest kind for engineering types to handle well. (I think it goes back to that shy thing we were talking about earlier.) However, it is also the most important communication skill to have. Face-to-face verbal communication is the best situation in which you can (a) make sure you are understood, and (b) make sure you understand.

WATCH FOR BODY LANGUAGE

Some say as much as 90 percent of what we communicate in a verbal conversation is in fact not verbal at all. If you really want to get deep into it, there are whole books on this topic that tell you the meaning of things like glancing right or left, up or down, and all sorts of looks. Most of the time, however, I believe that if you simply pay attention, you can get a lot out of how a person presents himself and the way he acts. You have been doing this from a very young age and it comes naturally if you give it a chance. Too often we get so rushed or distracted that we don't notice simple signs. For example, let's say that a person on your staff looks uncomfortable when he agrees to a deadline. If you don't notice and dig deeper, you could have a nasty surprise coming later.

CONSIDER WHO YOU ARE TALKING TO

The background of the person you are talking to should be considered as you communicate. Don't get caught in the trap of trying to explain details of quantum theory to the CEO who has an MBA. You should try to distill what you

are communicating to the points that matter to the person you are talking to. Take note of one important point, though: Don't ever treat the person like he is dumb! You can distill information without talking down to a person. If invited to, you can elaborate. You might be surprised at what your boss can understand—especially if he has read this book!

If you are dealing with a person from a different culture or who speaks a different first language, it helps to simplify your phrases to be sure you are understood. Don't get into vocabulary words that they might not know without being sure they understand what you are referring to. This particularly applies to words that have meaning only in your corporate culture. Everyone perceives the world through experiences they have based on the culture they come from. You don't need to speak *louder*. It doesn't help. Try to enunciate your words, though; if you are like me, you are probably carrying some hick accent that would cause you communication problems in your own native tongue.

SHOULD YOU GET ANGRY?

Sometimes getting angry is a correct response. There are occasions when that *is* what it takes for the person or people to whom you are talking to understand the seriousness of the point you are trying to communicate. You might have no other resort to get the point across. However, it should be rare, and if it is rare, it will carry a lot more weight than if you are someone who pops a cork every time something goes wrong.

REFLECTIVE LISTENING

A great way to improve verbal communication is to use a technique called *reflective listening*. The idea behind this type of communication is to further your understanding of what is being said by repeating it back to the person you are talking to. Take care, however, not to be annoying. No one likes a copy cat. The trick is to rephrase it in terms you understand and see if the other party agrees with you. This is particularly useful in dealing with persons from a different culture, say, a guy from engineering talking to a guy from management, for example.

READ

I believe that the single best way to improve your verbal skills is by reading. When you read, you experience how others communicate. It works with spy novels to white papers—the more you read, the better you will communicate with others. You will add to your vocabulary, you will understand cultural differences, and you will be able to order your thoughts in a way that is easier for others to understand.

Written

Whether emails, reports, or very official-looking documents, writing skills are extremely important in the field of engineering. Considering that nearly every engineer I have ever dealt with has had some issue or another with writing, I figure this is an often-overlooked skill.

PROOFREAD IT

Writing has one distinct advantage over verbal communication: You can look it over before you print it, send it, post it, or publish it. You should proof *every* written communication you create. The only question is how much. If it is short and you are going to follow up verbally, a quick glance will be enough. On the contrary, if it is going to be read by a superior or someone who might have reason to pick it apart, go over it a few times.

The most basic skill that I think should be used to proof writing is to read it out loud to yourself to see how it sounds. Don't forget to pause at commas and stop at periods, as you were told to in grade school. This technique will help you root out all sorts of odd-sounding phrases.

If you are particularly concerned, try it out on someone else and see whether they understand it. Make sure the person has a similar background to the audience for which the document is intended.

USE APPROPRIATE EMPHASIS

In written communication you lose the ability to create inflection with your voice, and you can't tie body language to what you are saying. This can be compensated for by emphasizing what you are saying with fonts, capitals, italics, boldface, bullet points, and numerous other options. If you SAY SOMETHING IN CAPS, you create the idea of yelling or raising your voice. **Boldface** words can imply importance, and italics can help you draw attention to *something in particular*.

There is, of course, a whole world of winks, smiles, and other punctuation types of communication out there, but I believe, although most will get the smile, the rest is a code that is known to only a few.

Note that I said *appropriate* emphasis. It is easy to get carried away. Don't cause death by bullets. Use too many bullet points and they no longer have meaning; too much boldface and it does no good; too many caps and people will think you are always angry. The trick is to be *skillful* in applying these skills.

USE VERBAL SKILLS IN WRITING

Some of these verbal skills are a great way to improve your writing skills. Things such as considering your audience and reflective listening (or reading/writing in this case) can help you understand and get your point across.

EMAIL SPECIFICS

Watch out for flame-mail. In the realm of email, it is very easy to be misunderstood, and people often respond with less tact than they might have in person. If you see a flame war starting, I think the best thing to do is talk to the person *in person*.

Watch the CC list on your emails, take care to whom you forward what, and always consider that what you have written can be easily forwarded to an unintended audience.

Get to the Point

Written and verbal communications have a few things in common. One of them is the importance of getting to the point. Use what you need to create understanding, but don't over-elaborate. If 10 words will do, don't use 100. Here are some other ways to get to the point.

USE ANALOGY

One of the most powerful methods of communication is the use of analogy. This works well for trying to explain a problem, concept, or theory. Analogy helps the person receiving information visualize what is being talked about. For example, analogy can help a person understand the details of a topic the same way that a telescope can help you see details of the moon. (Or maybe the apartment next door.)

There you go. I just used an analogy to explain analogy, and possibly a little humor too. *Analogy* is the art of comparing the new idea to something already known. It works very well.

SKETCH A PICTURE

You've all heard the phrase "A picture is worth 1000 words." Engineers typically get this; after all, they use schematics, which are simply pictures to represent ideas. In the world of email, however, we often ignore what we know so well. We will spend three paragraphs trying to explain what we want when a simple sketch will get the point across. Get yourself a scanner and use it to send a sketch with that email!

WATCH OUT FOR CORPORATE CULTURE CODE WORDS

Every conglomeration of people will develop code words to speed their communication. In a corporation, everyday words will take on completely different meanings. When you are dealing with people outside the company, be sure you don't assume that they know what you're talking about if you use a corporate word or phrase.

Thumb Rules

- Watch body language.
- Consider who you are talking to.
- Anger is sometimes appropriate, but it should be rare.
- Listen reflectively.
- Read.
- Proofread your written communication.
- Use appropriate emphasis.
- Use analogy.
- Sketch a picture.
- Explain corporate culture code words to those not of your culture.

ESPECIALLY FOR MANAGERS

Early in my career, I developed an outlook on management that can be summed up in a single phrase I wrote in my day planner in a boring meeting many years ago: "Management is an unnecessary evil." Years later I got my own "pointy hairs" and I discovered a few reasons for management to exist. (They might be good and true reasons or simply an attempt to justify my own existence; you will have to decide which.)

The Facilitator

To facilitate is to make easy or easier. Management should be a facilitator; it is up to you to create the environment in which an engineer can succeed. You need to get your engineers the tools they need to succeed. You need to help translate to your superiors the techno garble that engineers are so fond of. Most engineers just like to design stuff and really don't want to be in charge and manage things. They like to leave that up to you.

The Buffer

The best managers are a buffer between the top-level antics of illogical requests with unreasonable timelines and the real world of actual schedules and needs. They bring some order to the world of the engineer out of the chaos of pointy-hair decisions. This is something the engineer needs to be successful. Don't forget that in their world, it helps considerably if things make sense.

The Advocate

Good managers understand their employees and are their advocates. If an employee always has to beg for a raise, he will soon be looking elsewhere for a job. If he or she is a shooting star or even an average Joe, you will find that a reasonable show of appreciation raise-wise is much cheaper than hiring and training a new guy. It is not only right to be the advocate for your employees; it is good for the interests of the company as well. I get sick of hearing managers over-talk recognition and promotion as a way to make an employee happy in lieu of a raise. It is true that these things are nice, but that only matters if basic needs are being met—needs like food and shelter. If you are underpaying too much, no amount of awards will keep employees around.

The Gift of Focus

Good managers will develop the gift of focus. I find that this often comes naturally to an engineering type; they sometimes get so focused on the task at hand they might forget to even eat. For managers, though, their day is typically one of continuous and repetitive interruption. You can even get a false sense of accomplishment due to the fact that you are so busy being interrupted. To top it off, interruptions can spill over onto your engineering staff. Take caution that you don't find yourself constantly interrupting your engineers' focus. Be sure to find time to focus on your tasks; take advantage of that office door and close it on occasion to allow you to focus on things that need to get done. Keep meetings focused on the topic. Keep your team focused on your goals. Remember, the more difficult tasks require focus to complete. In today's information-rich world, focus can be hard to come by, so make it a priority in everything you do.

Understanding Engineers

Here are a couple of things you might or might not know about engineers but that will help you be a better manager of engineers.

WEASEL ROOM

Engineers need a little weasel room. Have you ever asked an engineer if he is 100% confident he has the solution? If you have, you were likely treated to a look of complete loss. It is not possible for an engineer to be 100% confident in anything. In this discipline you are constantly assaulted with the fact that you don't and can't know everything. You discover new ways for things to go wrong daily and are constantly working to fix and prevent them from happening. If an engineer gives you a range, take the conservative number for your estimate. Give the guy a little weasel room. Try to pin him down too hard and it could backfire on you.

THE ETERNAL OPTIMIST

I haven't met a good engineer yet who didn't regularly underestimate how long it takes to do something. This is simply a fact: Good engineers by nature are optimistic, and the really great engineers will push themselves so hard that they will meet the optimistic schedules they set for themselves. I heard a rule of thumb once about writing software that I have found to be true: Take the engineering estimate of time a job will take and multiply by three.

DESIRE TO GROW

The better you understand the "sparky" viewpoint, the more successful you will be at managing engineers. If you take this to the next level you can help your engineers take on more and more, literally turning an average Joe into a shooting star or even possibly rescuing someone from dudsville.[12] Most engineers want to grow and become better at what they do, but they need a little encouragement, a chance, and maybe a bit of a buffer against failure.

The Best Manager Is Right Most of the Time

Some time after I decided that management is an unnecessary evil and then later found some purpose in life after being inducted into management, I came up with a formula that describes a good manager. Remember, a manager spends nearly all his time making decisions—what tools to buy, what people to hire, what to do about a particular problem, what to eat for lunch,[13] and so on. How good he is depends on how often he is right. If he is right more often

[12]If you don't get the references to shooting stars, average Joes, and duds, you either skipped a few pages or have a serious memory problem.

[13]I believe there is a correlation here: The further up the corporate ladder a person climbs, the longer it takes him or her to decide what to have for lunch.

than not, the company makes money. If he is wrong too much, down the tubes it goes. So without further ado . . .

A good manager is right 51% of the time. A great manager is right 70%, 80%, even 90% of the time. If your decisions are right most of the time, you will succeed.

Remember this when faced with a decision. You don't have to be right all the time. Don't let indecision and too much worry prevent you from making a choice. Often that in itself can cause you to fail. Consider the situation, take action, and watch the results. Don't be afraid to admit you were wrong if you see you've made a mistake. Learn from the mistakes and don't make them again.

Finding the Shooting Star in a Forest of Average Joes

One of the most challenging things a manager has to do, after spending an hour or so with someone, is to decide whether that person would be a good employee and hire him or her. As we learned in previous discussions, you really want to stock your group with shooting stars, but how do you find them? How do you weed out the duds so that you aren't saddled with a problem down the road? Though there is no perfect solution, here are some key points to look for in a perspective engineer.

DRESS

Don't put a lot of value on how a person is dressed. Casual attire is the norm where I work, so unless someone comes in with serious hygiene problems, I don't chalk up any negative points. Once, however, a potential employee asked what the dress code was. His consideration impressed me. However, it is of minor importance. Our company is interested in results and product, neither of which is significantly affected by the dress of R&D employees.[14]

FUNDAMENTAL KNOWLEDGE

This is very important to me as a manager. There are some skills I don't want to have to teach you, skills I expect you to know for this position. A degree or some type of schooling tips the scale favorably, but I do not consider it a shoe-in. I have seen too many college graduates who got through school by the "assimilate and regurgitate" method. They passed all their tests with great grades, but they didn't focus on retaining the knowledge. I weed these people out with questions such as the one shown in Figure 7.1.

[14]Okay, that's not entirely true. When I think about it, the casual atmosphere we maintain makes us more productive, but that comes after the hiring decision, so it doesn't count.

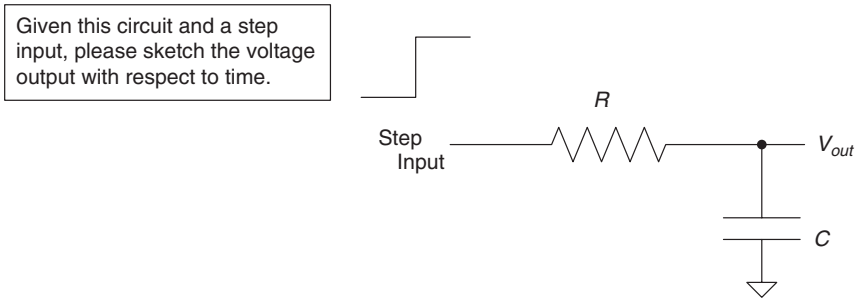


FIGURE 7.1
Standard "sparky" interview test question.

You might laugh, but being located right next to a major university with a reputation as a good engineering school, I constantly interview fresh graduates who should know this stuff. Fully half the applicants I see get this wrong! The basics are important. If you don't have them, you are just guessing when you design. Worse yet is if you think you know them and you really don't. After all the hammering on basics at the beginning of this book, I hope the importance of this concept is understood. I'd rather hire someone with the basics down pat and a 2.9 grade point average than the person that has a 4.0 and stumbles on basic understanding.

CAN YOU LEARN?

I have yet to see any employee get into a new job and not need to learn. Sometime during the interview, I will intentionally teach the candidate something new, and then hit the subject later in the interview, to see if he or she has picked it up. This ability to learn quickly and have it stick is important to the success of any engineering group. Technology will quickly outstrip those who can't learn.

ARE YOU WILLING TO LEARN?

You might think this was covered in the preceding question, but I consider it a separate point. I will often ask interviewees a question that I am fairly sure they do not know the answer to, simply to see how they respond. Do they try to BS their way through it? Or are they willing to admit that they don't know and ask for help? In the rapid design cycles of today, there isn't time to play games. That means "I don't know, but I will find out" is an appropriate answer. You can take this to a higher level, too. On call-back interviews, ask a question again that stumped them the first time to see if they were interested enough to figure out the answer.

PEOPLE SKILLS

Is there a job out there that requires zero contact with other human beings? I doubt it. The best engineering teams get along well, which is why I value people skills significantly. How do you handle pressure? Can you get along with people you don't care for? This is a fairly tough item to evaluate in an interview. I invite my leads to fire questions at the candidate and watch how he or she responds under pressure.

ATTITUDE/MOTIVATION

A positive attitude always impresses me. I quote my father: "Can't is a sucker too lazy[15] to try." I think it is important to believe that something can be done. Look for signs of giving up on a problem. Is the candidate persistent? Does he or she complain during the interview? Does he moan about his last job? I have seen all types. Whiners don't get hired.

COMMON SENSE

This is all about getting the job done in the least amount of time. Too often a person can be "book smart" but not be able to apply what he or she has learned. If you don't have common sense you will struggle with applying the knowledge you have. Here is a brain teaser[16] I often use to determine a person's level of common sense:

> You are standing in a room with two strings hanging from a high ceiling. If you grab just one string and walk to the other, the second string is several feet out of reach (because it is hanging straight down). Your task is to tie the two strings together. You have just three things to perform this task: a book of matches, a single square of toilet paper, and a screwdriver. How do you tie the strings together?

IN GENERAL

Should you be looking for the person who can do differential equations in his head? I don't think so. I will buy Mathcad for that. I want to know if the candidate has the fundamentals and if he can and will learn the rest.

---

[15]Can laziness actually be an asset? If it motivates creativity, it can. Remember, if you give the hardest job to the laziest man, he will find the easiest way to do it. So I guess you could look for motivated lazy people, if that makes sense at all.

[16]I'm not sure I want to reveal the answer. People who are smart but with less common sense will overthink it too much. (You wouldn't believe some of the answers I've heard!) If you have a high common-sense quotient, you will get the solution in a second and wonder why it was so simple. Email me for a hint if you are struggling!

Remember, great managers are rare; mediocre managers are commonplace. You don't have to be a great manager for a company to be successful. Why stop there, though? Great managers are huge assets to any company; great managers can turn average Joes into shooting stars can make incredible things happen.

Being a great manager isn't all that hard. Listen, look, and learn until you are right most of the time. Then you won't be unnecessary or evil!

Thumb Rules

- Management is an unnecessary evil?
- Be a facilitator.
- Be the buffer.
- Be an advocate.
- Develop the gift of focus.
- Understand your engineers.
- Be right most of the time.
- Hire shooting stars; make shooting stars from average Joes.
- Don't be unnecessary.
- Don't be evil.

ESPECIALLY FOR EMPLOYEES

As an employee your motivation, like the pointy-hair boss you work for, eventually boils down to money. You want to do a job and get paid for it. True, job satisfaction is important, but that comes way second to the need to buy food to eat and have a dry place to sleep. This means that an employee needs to know two things: how to get a job and how to keep a job. This chapter is a guideline to those things.

How to Get a Job

It all starts with the interview. Having interviewed more engineers than I care to remember, I have compiled seven definite no-no's[17] extracted from real

[17]Come to think of it, these don'ts aren't just for engineering interviews; you could make a pretty good case for each one as a rule of thumb in almost any job interview.

interviews. Giggle, laugh, and snicker if you will, but please do not try these during your next interview. The people described in the following paragraphs are professionals.

DON'T BE CONDESCENDING

Be careful how you come across to your potential employer. One candidate I interviewed seemed to really disdain coming to us for a job. It was as though he would work for us if he really had to, but he sure wasn't going to like it. The "you don't have anything to teach me" vibe was very strong. Being an engineer who believes the ratio of what we know to what we don't know is extremely small, I have a tough time with that. This is especially disconcerting when some simple circuit diagrams are requested and you get the response, "Everyone knows that," a little hand waving, and then nothing is written down and no answer is forthcoming. I immediately think you don't actually know it, and this is all an act to cover up the lack of knowledge.

DON'T WORRY ABOUT SAYING "I DON'T KNOW"

The stress of an interview might make it the toughest place to say "I don't know," but that is not a bad answer. Especially if you follow up with, "I'll find out, though." One of the best impressions I had from a potential employee was when he sent me an email afterward that explained the answer to one of our questions in the interview that he didn't know at the time. The fact that he took the effort to look it up showed perseverance and a desire to learn. That alone will many times make up for a current lack in knowledge.

DON'T LOSE YOUR COOL

One person I interviewed was clearly thrown a bit off balance by some of the questions I asked. What really put marks in the cons column was when he got so upset trying to solve the problem that he threw down his pencil and repeatedly smacked the table. Our work environment can be much more stressful than an interview; I really don't want to worry about someone going mental on the job.

DON'T GIVE UP EASILY

If you don't know the answer to a particular problem, try to figure it out if you can. I will often ask questions that I know the candidate won't know, just to see how he or she handles it. Someone who takes one look and walks away has never impressed me. Remember, while someone is standing there saying it can't be done, someone else is out there doing it.

DON'T BE AFRAID TO ASK QUESTIONS

Along with the preceding point, you are not expected to know it all. If a person asks a question about a particular task or problem I've given him or her in an interview, it usually shows that a person who doesn't know is willing to find out. That is a very important trait in the engineering world. Also use the interview as a chance to find out about your respective workplace.

DON'T LAY YOUR HEAD ON THE TABLE

Yep, it really happened and I have witnesses to prove it. This potential employee laid his head on the table several times during the interview. I couldn't figure out whether he was tired or just listening for some type of table vibration that might indicate how well the interview was going. This would never be my only reason for not hiring someone. (I get some of my best ideas in that twilight between almost asleep and almost awake.) However, this was coupled with some other blatant problems. I just knew it wouldn't work. Let's just say this particular interviewee will have plenty of time to nap now.

DON'T CALL YOURSELF STUPID

I wouldn't have believed it if it hadn't happened to me. One applicant we had got a little flustered with a couple of basic questions, but that wasn't what did him in. The first time he said "Man, I am stupid," I didn't think much of it; however, as the interview wore on, I heard, "Oh, I'm an idiot" and "I am soooo stupid" probably a dozen times or more. By the end of the interview, I was sure of one thing: I definitely didn't want to hire an idiot, especially one so stupid.

A FINAL THOUGHT

There are a lot of guides out there on getting an interview and getting through an interview. They are even a bit more conventional than my seven don'ts. It can't hurt to study up on some of these pointers. I also think it helps to know a bit about the company you are interviewing with. Take advantage of today's ability to look up anything on the Internet. It will help you decide where you want to be, and it also doesn't hurt to have a little background before going into an interview.

How to Keep a Job

When the ax falls, will you be the one to get chopped? How do you increase your stability in a given company? What makes an employer keep one person and let another go? Here are five key areas that can give you a little more security in this layoff-prone world—things you can do besides simply being good at your job.

VALUE

Here's a Thumb Rule: *Companies exist to make money*. Even nonprofit companies need to bring in money to cover their salaries and expenses. When your employers start reviewing you and your coworkers, you need to realize that this is foremost in their mind.

This is the question the manager must ask himself: If I had to start all over with just one employee, who would it be? Or in other words, who would most likely make this company a success? In my analysis, this person is the "shooting star." He or she works hard, has great talent, can handle pressure, and works well with others. If you ask him for something you get it. You don't have to keep checking up on him. You know she is going places. He very directly affects the profitability of the company.

Therefore, you must remember that your total value is of top importance. What if you add value, though, and no one notices? This can happen, especially in larger companies. My answer is this: It is not bad to toot your own horn a bit. A good way to do this, both for you and your employer, is to do a regular self-evaluation. List the things you accomplished last year and compare them to what you did this year. Do you show improvement? If not, commit yourself until you do. Then give that to your boss. He'll appreciate that you look at yourself critically and it's a good chance for him to see what you have done for the company.

POSITION

Repeat the thumb rule we just learned: Companies exist to make money. They don't do that without a product. So the most important job you can have is one that is directly related to the product. Don't get stuck in a one-off job. What is a one-off job, you ask? A one-off job is one you can eliminate and still sell product. It is one level removed from delivering a product to the customer. The ISO 9000 "Corporate Coordinator" might sound like a pretty neat title, but when you get right down to it, the company could do without it. If you find yourself in a one-off job, it's time to start looking for a transfer.

LOYALTY

It's human nature to complain. Because of that, an easy yet subtle trap to fall into is right by the water cooler. In this trap you discuss the latest smack about the boss. Every leader I have ever met appreciates loyalty. If you succumb to spreading rumors, whether true or false, you put yourself on shaky ground. I

am not saying the pointy hairs don't make mistakes. In fact, I believe that a manager only needs to be right 51% of the time to be successful, as you already know. So remember this: They might have their faults, but so do you. If you have a serious issue with your boss that you can't overlook and can't help talking about, you'd better start looking for a new job, because in today's market, you soon will be.

EFFORT

Effort is important for two reasons. First, a great effort can compensate for a lack of skill. Remember that the guy who tinkers in the lab for hours on end can get to the finish line faster than the brilliant engineer who spent the morning surfing the 'Net. It's all about getting to the market the fastest these days. It is the entire reason that MAMA[18] exists. All the pointy hairs want to do is to deliver product, make the sale—in general, to do business. So a supreme effort is usually noticed. Remember, the same rumor mill you should avoid yourself can have a tremendous effect on you. You can be known for hard work, or you can be known as a slacker. The choice is up to you.

IF THE WORST HAPPENS

It is possible that no matter what you do, you still get laid off. At times when a company has to cut deep, there is nothing that can be done. I suggest you take this as best you can and leave on a good note. If things pick up again, it is a lot easier for a boss to rehire someone he knows will do a good job rather than any Joe off the street. So don't burn any bridges.

A Final, Final Thought

By no means do I consider this list comprehensive. There can definitely be more to it. People skills, attitude, and other things are considered by an employer when making this tough decision. To make it worse, the world is not all sugar and spice. There are sadistic pointy hairs out there who give the rest of us a bad name (I just hope they are the exception, not the rule). If you have one of those, don't complain, just start looking.

Remember, dealing with people is not a very exact science. There is no Ohm's Law for corporate culture. These are things that I have found that generally work. You can sum it all up by referring to the different types of employees we

[18]Look it up in the appendix; I promise you will find some of the more entertaining parts of the book there.

have previously discussed: the shooting stars, average Joes, and duds. When it comes time for layoffs, you don't want to be a dud, and if you can, try to be a shooting star.

Thumb Rules

- Avoid the seven interview don'ts.
- Companies exist to make money.
- Companies exist to make money (*duplicated to indicate importance*).
- Take care of the five key areas.
- It isn't a perfect science.
- Don't be a dud.
- Be a shooting star.

HOW TO MAKE A GREAT PRODUCT

The Slinky, Legos, the PC, Silly Putty, weed eaters, Velcro, cell phones, DVDs, pet rocks, and the microwave—the list of killer products seems endless. How do you go about designing a great product? What makes a product successful? Believe you me, the list of great ideas that never went anywhere is much larger than the list of things that made it! For those of you with a more entrepreneurial spirit, here some pointers on how great products come into existence.

The Idea

Usually the core of a great product addresses a need or desire. The more people who share that need or desire, the more success potential an idea has. Here is a real live example. My car windows are always frosty during the winter here in Logan, Utah. I don't have the patience to start my car early and wait for the defroster to clean the window, so I scrape. Scraping is not much fun, and last year I had a great idea for an invention. Why not put a heater in the windshield washer fluid so that I could have a quickly defrosted window without having to scrape? I am sure there are other people like me who would want this product.

Let's evaluate this idea for a second: first, the buyer of such a product would need to own a car. That limits the primary market to Canada, the United States, and Europe. Then it has to get cold enough to frost your windows where you live. There goes half the United States. Next, to be like me, you can't afford to

park your car in a garage, and that eliminates a bunch of people. I figure that Canadians like to scrape, knocking off another large part of the market. So will this idea make me a million? Probably not, but if I worked hard enough at it, it might generate a decent income for a while.

Compare that to the market of the weed eater (string trimmer, to be more correct). When George C. Ballas stuck some twine in an old tin can and spun it on his electric drill, he was addressing a need that many a man felt. Not only did it chop those pesky weeds, it involved a motor as well and, oh, the power rush!

His market was anyone who had ever wished for an easier way to trim those hard-to-reach places in his lawn. To top it off, it also stroked the male ego. I think it had a larger success potential than my defroster idea, don't you? Notice that I said *potential*. A lot more than potential is needed to make a product a success.

Design

The product needs to work well. This means that the design needs to do what the customer expects from it. If everyone sends the product back, it won't be a success for long. There is one all-too-evident exception to this rule: software! Sometimes people will deal with glaring product faults (also known as GPFs) if that is the only game in town. It is that or you really need a particular feature and are willing to deal with the bugs.[19] It bothers me though that you can't send it back because you clicked "I accept" on the 40-page EULA that no one reads, which prohibited you from even taking Bill's name in vain, let alone returning a product. But . . . here I am using that same popular word-processing program because of the features I like.[20]

It also needs to look good. Ever since the 1950s, industrial designers have convinced the consumer that you can have a functional product that looks good as well. There are successful ugly products out there, but if they looked good they would be even more successful. Have you ever said, "That's a sweet little package," in reference to something other than the opposite sex?[21]

---

[19]Okay, I'm slamming on a certain large software company here. To be fair I will say that my more current versions of this software are significantly better than the versions I learned on many years ago. The best thing that ever happened to software was the ability to update it. Yes, they can lock out the hackers and pirates who believe that all IP should be free, but for me and most of the world it is nice to get a free update that fixes that pesky bug.

[20]I still hate EULAs though. Software, unlike every other consumer product, carries no responsibility for any damage it may cause. Someday this may change, but it would likely drive the cost of software up and I'm a cheapskate, so I guess I'd better stop complaining.

[21]Possibly a bad analogy, since I have heard an engineer say exactly those words in reference to an IC, but you get the point.

Timing

Ahhhh, timing … it is as important in launching a product as it is in telling a joke. I don't think the weed eater would have sold before America moved into the suburbs and the lawn wars began. The Slinky wouldn't have made it very far if it came after the Nintendo. A company I worked for had an idea that changed our marketplace. It was featured on some 30 different news channels and became a raging success. It didn't stick till the third time it was tried. The first two times were utter failures. It needed the Internet as a global community to be a success. The first two times it was tried, the global data community just wasn't there to give it the buzz it needed. Timing is important.

Funding

It takes a million to make a million, right? This is usually the case unless you listen to late-night TV. Now, if you believe that stuff, I have a book on how to get a perfect stranger to give you 50 bucks. I will sell it to you for only $47.95 plus shipping and handling.

I think that funding is one of the things that stop more great products from coming into being than most other reasons combined. You have to take some type of financial risk. One way is the OPM method: use Other People's Money. Unfortunately, it takes a smooth talker to get other people to part with their money, so you might have to run up your credit card or go deep into your savings. There are many ways to get the money, but it will cost money to get your idea to market. You will have to take some risk at some point to make it happen.

Marketing

You *have* to sell your product. No one will buy a product that isn't sold. That takes marketing. 'Nuff said.

Okay, maybe not. I used to think this was pretty obvious, but when I started a business helping people get stuff to market, I found that marketing is often the most ignored part of getting a business off the ground.

If you build a better mouse trap, the world will not pound a path to your door without an infomercial, a store, or a way to know it exists. You will need to be a salesman of some type to get your idea off the ground. If someone doesn't buy it, you don't have a product—just another idea that didn't go anywhere. The patent libraries are full of mouse traps that you can't buy anywhere.

Making a Great Product Summary

So, will you be the next Bill Gates? Just think of a product that everyone wants. Get a couple of rich relatives to put in a good word and pitch in a few bucks, and who knows. If your timing is right, it just might happen. If not, I still have that book for sale . . .

Thumb Rules

- Have a good idea (this is the easy part).
- Consider the market potential.
- Your product needs to work well.
- Looking good helps.
- Timing is everything.
- It won't happen without risk.
- You need to sell it.

Glossary

One of my favorite areas in a physics book is the inside cover. It is where all the good stuff is distilled into the fundamentals. I couldn't call this book complete without creating a similar section.

The following are some terms that you might or might not know—words that are often used in the realm of electronics but that typically cause a look of confusion on any nonengineer who accidentally overhears a conversation between a couple of sparkies. These words constitute a secret code, usually short to be more efficient and sometimes intended to baffle the boss, or at least make him wonder what you are really talking about. They have been selected at will based on looking at my own secret decoder ring and deciding what was okay to reveal without risking lynching by my fellow engineers.

AC *Alternating current*, or current flow that switches back and forth. This is the type of power that comes in on the line to your house and is available at a common outlet.

Back EMF *EMF* means *electromotive force*, which is used to describe the voltage generated when you spin the armature of a *DC* permanent magnet (PM) motor. The term is also used to describe the voltage generated at the connections of an inductor when you stop pushing current through it and the magnetic field collapses. Since they are both voltages caused by a changing magnetic field, it makes some sense.

Bias A widely used term in electronics. *Bias* can refer to the voltage applied to a circuit. For example, a *DC* bias or offset is a way of shifting an *AC* signal from one level to another, such as biasing a circuit or component to a level where you get a predictable behavior. You can bias the input of a transistor, for instance.

BS Come on, everyone knows what *BS* means!

BTW *By the way*; the only reason I need to define this is for old farts like me who were raised without a cell phone and text messaging!

Bulk cap A large value capacitor, usually 1 µf or bigger, commonly 100 µf to 0.1 f. Usually an electrolytic cap, not typically good at fast frequencies but has plenty of current capability.

Cap *Capacitor*, a plate-like unit with a space of something that won't conduct electricity between the plates. A cap has the capacity to store energy in the form of an electric field.

Chip Slang for *IC*. You will often hear engineers refer to ICs as chips. It doesn't always mean they are hungry for lunch!

Current Describes the movement of electrons, commonly thought of as a flow. In the water analogy, this is the amount of water moving. Amp is the basic unit of current in an electrical circuit. Common symbols are *I* and, less often, *A*.

DC *Direct current*, or current flow that goes in only one direction. This is the type of power that comes from a battery. It is the type of power computers and most electronics use internally in their circuits.

DCPM Short for *direct current permanent magnet* motor. These little guys are everywhere.

DMM *Damn meter won't measure*; a cuss phrase often let loose when an engineer is yet to discover that the fuse is blown in his digital multimeter. Usually precedes stalking off to the lab to find a screwdriver since you have to tear the whole meter apart just to replace a fuse.

Drain Usually this is the connection on a device that "drains" current from whatever it is hooked up to.

Drive *To drive a part* means to apply current and voltage to make the part do what you want. You drive a *load*. If asked what a ____ is capable of driving, it means how much can it *sink* and *source*.

Duty cycle A percentage of on-time versus off-time—how much time the component is on duty, so to speak. If a motor has a 30% duty cycle, that means it is being used 30% of the time; the other 70% of the time it is off.

EPROM Way back when our PROMs only had one E, you had to erase them with UV light. Oh yeah, it means *erasable programmable read-only memory*. Does that mean EPROMs technically were "easy to sunburn"?

EMI *Electromagnetic interference* is anything and everything that interferes with an electric or electronic circuit. It is sometimes attributed to supernatural causes by superstitious engineers.

EULA *Everyone is unable to take legal action* if this product destroys your data. If you have never agreed to a EULA and you own this book, well, wow. I am left at a complete loss trying to come up with a quirky remark.

FAE *Fairly astute engineer*. Most FAEs I have met are pretty smart, or I am just jealous that they got the easy job? I'm not really sure. Oh yeah, it also means *field application engineer.*

Flame mail An email message that is sent with the intent to harm, not actually communicate.

Flux Flux, or resin, is an acid either applied separately or in the core of the solder. When heated, it cleans the joint to help the solder stick better.

Forward bias Refers to the biasing of a diode; when forward-biased, a diode passes current.

Freewheel diode A reverse-biased diode hooked up in parallel with a motor. It is there to capture the inductive current generated as the magnetic field collapses.

Gate This means a couple of slightly different things: a logic part, NAND gate, NOR gate, etc., or a connection on a FET that controls the current flow from drain to source. Note that it isn't all that different from how a "gate" can keep or let out sheep in a coral—that is if you can compare sheep to electrons. Now there is an analogy that would be fun to explore.

Gnd, *Vss* The voltage reference point. Usually you connect one lead of a measuring instrument to this point. It is also the place all the current returns to (conventional flow again) that comes from *Vcc*. In electron flow terms, it is the point that spews forth electrons.

Grok Martian term in the book *Stranger in a Strange Land* by Robert Heinlein. It means *to understand completely, in the most intimate way*.

Ground Often used interchangeably with circuit gnd, *ground* should be thought of differently. Ground is the dirt under your feet into which you drive a big stake and hook it up to the exposed metal (and sometimes the *gnd*) of your circuit. This is done for safety reasons.

HW Abbreviation for *hardware*.

IC *Integrated circuit*. A device that is made up of a combination of diodes and transistors and other basic parts etched into a silicon base; it's used to make things as simple as switches and as complex as the Intel Pen-*way-cooler-than-the-last-chip*-tium in your PC.

Impedance Seen as a *Z* in many equations. Think of this as resistance that takes frequency into account. Used in conjunction with inductors and capacitors.

Inductor A coil of wire at its most fundamental; it can store energy in the form of a magnetic field. When a magnetic field changes, it induces current to flow in a wire. The coils concentrate the magnetic field.

Iron Soldering iron used to create solder junctions. No, you don't want to iron your shirt with this device!

ISA *Intuitive signal analysis*—the first acronym of my own invention. I figure if I ever want to be a famous engineering writer, I'd better have one or two acronyms to my name.

Junction The place at which two semiconductors come together.

Ladder logic A type of programming method or language; its name comes from the ladder-like appearance of the diagram used to describe the program.

Lead A pin on an electronic part, such as an IC, used to connect the part to the PCB.

Leaky cap An imperfect capacitor that allows some amount of DC current to pass.

Linear A term often used in conjunction with supply or control. A *linear* control is one that controls voltage to a part continuously. The part controlling this will dissipate energy based on the voltage across it and the current through it. It is typically an inefficient way to drive a *load*, since the power that is not used is turned into heat.

Load Something that takes power, needing both current and voltage, to drive. A resistor that returns current from *Vcc* to *gnd* is a load.

Magic smoke The stuff inside all ICs that makes them work. You don't want to let it out!

MAMA *Management always chasing the market around.* My own personal acronym. If you want to be successful in the world of engineering, you have to invent an acronym or two. Chalk up another one for me!

MCU Microcontroller, which is like a CPU but less powerful, with more stuff built in.

NO, NC Pronounced *nnnn ohhh* and *nnn seee*. A cryptic abbreviation for the typical state of a switch or relay connection. See, even in engineering, NO doesn't always mean no.

OPM *Other people's money*; it's always more fun to play around with other people's money than with your own.

OS Operating system.

OTP *One-time programmable*. Before Flash became the memory of choice in embedded micros, one chance was all you got. There are still a few OTPs out there, but you are probably in some really high volumes if you're using these It's likely you are into masked parts as well.

Pad Not the place where you hang out! It's the point on a PCB of bare copper where the leads of a part are connected by solder to a trace.

PCB or PWB *Printed circuit board* or *printed wiring board*. A composite material, usually stiff like a board, on which a circuit is laid out, creating connections between components.

PDA *Pretty dumb assistant*. I'd trade my PDA for a real live flesh-and-blood assistant any day!

PLD *Programmable logic device*. Take a whole bunch of memory cells, a slew of logic gates, a bunch of multiplexers, and a way to configure it all, and then cram everything into a single IC. At the end of all this, you get a product that can do a whole bunch of state machine and logic stuff. You can even make MCUs out of them, as in sister products such as the FPGA.

PM Permanent magnet.

Pointy hair We have Scott Adams to thank for this unique term, which we can now use to refer to our bosses.

Power The combination of voltage and current. This is what turns the lights on in your house. The unit for power is the watt. The common symbol is *W*. Watts can be converted to horsepower (HP); it takes 746W to make 1 HP. Another symbol you might see that is loosely related to watts is VA, or volt amps. The symbol is generally used in power supply systems to refer to AC power; it is equivalent to watts only when the current and voltage match phases.

Power component A term commonly used to refer to parts that handle a large amount of current or high voltage. Of course, the words *large* and *high* are relative. It means a current large enough so that you need to worry about things like heat and voltage, and high enough so that it will do more than tickle a little if you touch it.

Power device A common term used to refer to semiconductor devices, such as FETs and transistors, that take a small low-power input signal and amplify it into a high-power signal. Power devices usually need to be meticulously handled in your design to avoid overheating. They often have a surface that is designed to be coupled to a heat sink to manage the power dissipated as they operate.

Pull-up A resistor from an input line to *Vcc*. In the absence of any other current flow, it "pulls" the voltage at that node to Vcc.

Pull-down A resistor from an input line to gnd. In the absence of any other current flow, it "pulls" the voltage at that node to gnd.

PWM *Pulse width modulation*. A digital method of controlling a voltage level. The percentage of time-on versus time-off determines the amount of power applied to the load.

R Pronounced *arrrr*, as in "What is the arrr of that puppy?"; it means *resistance*—something that resists the flow of current proportional to the voltage. It is the *R* in Ohm's Law.

Rail The voltage limit to which an output can swing. The top rail is the highest positive voltage it can get to, and the bottom rail is the lowest voltage it can get to. This is not necessarily the same as the power supply. Some devices cannot get the output to reach *Vcc* or *gnd* in the circuit. When the output is at these limits, it is common to say it is "railed."

RC *Radio control*. A fun hobby that you can dump a lot of money into. Also means *resistor/capacitor circuit*.

Rectify *Rectify* or *rectification* is the process of turning AC power into DC power.

Reverse bias A specific case of biasing, usually referring to a diode. When a diode (or diode-type junction in a component) is reverse-biased, the diode blocks current flow.

RSP *Really smart person*. I love to talk to really smart people; that is, when I can understand what they are saying!

Sink No, not the kitchen sink, but it does act a little like a drain; generally used in a phrase such as "How much can that sink?" It means how much current is capable of going into ground through that part.

SNL *Saturday Night Live*. I'm probably dating myself, but I remember when this was on a "real" network, not just syndicated TV.

Solder A material used to make electrical connections. It is heated to create that connection.

Source A term often used in a phrase such as "How much can that source?" It means how much current is capable of coming out of that part. Both *sink* and *source* assume conventional current flow terminology from positive to negative.

Sparky A widely used slang term to refer to an electrical engineer, at least in the world of Darren. (We tried to assign the term "wrench" to the MEs, but it just doesn't have the same ring to it.)

State machine A computing device that looks at the state of the inputs to determine the output. More complex forms of this device feedback outputs to the input and/or maintain memory of certain inputs.

SW Abbreviation for *software*.

Switcher A cousin to the linear control or supply. The switching control is digital in nature. Somewhere in the system is a switch that turns on and off cycling power to the load. The amount of time-on versus time-off is called the *duty cycle*; it is defined as a percentage. Often there is an inductive or capacitive component in or attached to the *load* that filters the frequency of the switching device to smooth out the voltage or current to the load.

Switch mode The digital control of a device such as a transistor or FET, for example. The part is either turned all the way on or off, like a switch—hence, *switch mode control*. Using a device like this in applications, such as a switching power supply, helps make them more efficient because less heat is created when a part is not in the linear region of operation.

Threshold In electronics, a voltage level that, when crossed, changes the output state of a logic circuit from 1 to 0, or vice versa.

Tinning Refers to applying *solder* to the tip of an iron or to a wire to help heat transfer.

Trace The little green lines you see on a *PCB*. They are made of copper and are the wires that connect the parts. *Trace* can also refer to a method of troubleshooting software.

Vcc, Vdd The voltage source in the circuit. In conventional flow terms, it is the place all the positive holes come from. In electron flow terms, it is the place all the electrons try to get to.

Via A hole in a *PCB* that on some PCBs is coated with copper. It is used for two reasons: either to create a connection between a top trace and a bottom trace or to create a hole in which a part lead can be inserted and be soldered to the PCB.

Voltage The potential of the available electrons. Using the water analogy, this is the pressure the current is under to move. The unit for voltage is the *volt*. Common symbols are *V* and *E*.

Voltage drop The voltage measured across a component, such as a resistor. Not a "drop" in a bucket or anything like that; it's simply techno-speak indicating the difference in voltage as *measured* from one side of a component to another. (Since what you measure is relative, you can always switch the meter leads to make it look like a "drop" in voltage.) If a voltage drop increases or decreases, this means the absolute value or magnitude of the *change in voltage* across the component is increasing or decreasing.